Department of Economics, Finance and Accounting
NUI Maynooth
Co. Kildare
Ireland

Studies of Inflation and Forecasting

Dissertation presented in order to obtain the title of
Doctor of Philosophy (Economics)

for

Colin Bermingham

under the direction of Professor Maurice Roche

Other Members of Supervision Committee:
Simon Broome
Tom Flavin

Studies of Inflation and Forecasting

Acknowledgments

I would like to thank my supervisors for the support, comments, suggestions and overall guidance during the entire process.

The last two chapters of the thesis have been jointly written with Derry 'O Brien and Antonello D'Agostino, who I would like to thank.

I would like to thank Kieran McQuinn, Gerard O' Reilly and Karl Whelan for helpful comments and suggestions.

I am grateful to the faculty, the secretarial staff and the doctoral students of the NUI Maynooth for their support and suggestions.

The thesis has been written while working at the Research Department of the Central Bank of Ireland, which provided funding and support for this course of study. The hospitality shown to me is gratefully acknowledged.

To my parents

Contents

Introduction

This dissertation contains five research papers in the area of applied econometrics. The two broad themes of the research are inflation and forecasting. The first two papers examine the topic of core inflation. In these papers, I construct core inflation measures and rank them according to various criteria, including their ability to forecast the headline inflation rate. The third and fourth papers deal with issues in relation energy prices. The first energy paper examines the degree to which oil price movements can be used to forecast consumer prices for energy. The second paper tests for asymmetry in the response of petrol and diesel prices to changes in international oil prices. The fifth paper considers the issue of forecast aggregation and examines whether it is better to forecast inflation directly or instead forecast its components and then sum those component forecasts. All five papers consider various aspects of inflation and, with the exception of the paper on asymmetric pricing, the other four papers all involve constructing inflation forecasts. These are the two unifying themes of the dissertation.

The first paper is titled "How useful is core inflation for forecasting headline inflation?". It constructs core inflation estimates for Ireland and ranks them according to their ability to forecast the headline inflation rate.[1] The paper is the first to construct an estimate of core inflation for Ireland using a Structural Vector AutoRegression (SVAR). The system contains three variables and identification is achieved through the use of long-run restrictions. The key restriction is the core inflation shock is output neutral in the long-run, which is consistent with a vertical long-run Philips curve. Alternative core inflation rates are constructed using the Hodrick-Prescott (HP) filter, trimmed means and the inflation rate excluding energy is also considered. The benchmark forecast in the paper is based on an AutoRegressive Integrated Moving Average (ARIMA) model. The ARIMA is found to have good short term forecasts but the SVAR provides the best forecast over most horizons.

The second paper is called "A critical Assessment of Existing Estimates of Core Inflation". The paper continues the topic of core inflation but examines the properties of core inflation estimates in much greater detail than the first. This paper focuses entirely on US data. The paper extends the literature in a number of ways. There is an existing literature of papers which compare core inflation rates but this is the most exhaustive in terms of the range of core inflation estimators included. The paper examines the core inflation rates according to a range of different criteria, including the forecasting ability of the core rates

[1] The terms "headline inflation", "actual inflation" and "overall inflation" are used interchangeably in the research to denote the inflation rate of the national consumer price index or consumption deflator.

and their ability to track trend inflation. The paper highlights shortcomings in the way that these tests are applied in the literature and corrects for these problems by applying the tests more rigorously. It is found that the core rates are no more useful than simple benchmarks when it comes to tracking trend inflation or forecasting inflation. This is a novel result and represents a serious limitation of core inflation rates from a policy perspective. Although both papers on core inflation conduct exercises which are similar in spirit, the conclusions of the second paper are quite different in the sense that core inflation rates are not found to be useful. There are a couple of reasons underlying this difference. Obviously, the data used in one study is Irish while the other relates to the US so one would not necessarily expect the same results. In addition, it may well be the case that the benchmark forecast is more easily beaten in the Irish case, as ARIMA models can have poor long-run forecasts. Finally, the US paper is a more substantive contribution, as it highlights the problems with the evaluation of core inflation rates in the literature. By conducting these tests more rigorously, it goes against the results of the general literature.

The title of the third paper is "Quantifying the impact of oil prices on Inflation". This paper was written on the back of a period of extraordinary volatility in international oil prices, during which oil prices increased from around \$30 per barrel in 2003 to \$125 per barrel in 2008. This spike in oil prices did not have the same impact on inflation as the oil price spikes in the seventies, suggesting that our econometric estimates of the impact of oil prices on inflation need to be revisited. The ability of oil prices to help forecast consumer energy prices is examined rather than focusing on the forecasts of overall inflation. The overall inflation rate is influenced by many other factors, including the large increase in agricultural commodity prices which occurred in tandem with the oil price increase, so to focus on the energy component provides a cleaner estimate. Regulated prices are stripped out of the energy mix so that only market driven prices are included when measuring the responsiveness of consumer energy prices to international oil movements. The paper shows that it is possible to significantly improve upon the benchmark forecast. Given the well known difficulties with inflation forecasting, this demonstrates the merit of focusing on the energy component of inflation rather than overall inflation.

The fourth paper is titled "Testing for asymmetric pricing behaviour in the Irish and UK petrol and diesel markets" and my co-author on this paper is Derry O' Brien. This paper empirically tests whether Irish and UK petrol and diesel markets are characterised by asymmetric pricing behaviour. The econometric assessment uses threshold autoregressive models and a dataset of monthly refined oil and retail prices covering the period 1994 to mid-2009. In addition to providing an appraisal of the existence of asymmetry in the Irish and UK markets, the paper provides an important methodological contribution. Tests of asymmetry in the literature normally partition the sample into periods of falling and rising international oil prices. This fails to account for price pressures coming from the equilibrium error of the cointegration relationship. In particular, the possibility of conflicting price pressures arising from short-run dynamics in retail prices and responses to disequilibrium errors needs to be explicitly modelled. We are the first to take this issue into account in an econometric model and we highlight the importance of this distinction. In terms of the asymmetric behaviour of these markets, the paper finds no evidence to support the "rockets and feathers" hypothesis that prices rise faster than they fall in response to changes in the value of international oil prices.

The final paper in the dissertation is called "Understanding and Forecasting Aggregate and Disaggregate Price Dynamics" and this paper is co-authored with Antonello D'Agostino. The issue of forecast aggregation is to determine whether it is better to forecast a series directly or instead construct forecasts of its components and then sum these component forecasts. Notwithstanding some underlying theoretical results, it is generally accepted that forecast aggregation is an empirical issue. Empirical results in the literature often go unexplained. This leaves forecasters in the dark when confronted with the option of forecast aggregation. We take our empirical exercise a step further by considering the underlying issues in more detail. We analyse two price datasets, one for the United States and one for the Euro Area, which have distinctive dynamics and provide a guide to model choice. We also consider multiple levels of aggregation for each dataset. The models include an autoregressive model, a factor augmented autoregressive model, a large Bayesian VAR and a time-varying model with stochastic volatility. We find that once the appropriate model has been found, forecast aggregation can significantly improve forecast performance. These results are robust to the choice of data transformation. This provides a significant endorsement of the forecast aggregation approach and the results in the paper highlight the interplay between model choice and the level of data aggregation.

Chapter 1

How Useful is Core Inflation for Forecasting Irish Headline Inflation?

The paper constructs various core inflation measures. These include various trimmed means using highly disaggregated data and a structural VAR estimate of core inflation for Ireland. The ability of these core inflation measures to forecast future headline inflation is compared using a simple regression model. An ARIMA model fitted to the headline inflation rate is used to construct the benchmark forecast. The forecasts from the ARIMA model are most accurate over short time horizons for both monthly and quarterly data. The structural VAR based estimate is most accurate over longer time horizons.

1.1 Introduction

During the past decade, there has been a revival of interest in the topic of core inflation as more central banks engage in inflation targeting. Specific inflation targets have been adopted by central banks in several countries including Australia, Canada, Finland, New Zealand, Spain, Sweden and the United Kingdom. The European Central Bank has also committed to maintaining the inflation rate in the euro area below two per cent. Core inflation can be used as an indicator of future trends in headline inflation. Consequently, it provides a tool in the formulation of monetary policy, particularly for central banks that engage in inflation targeting.

Core inflation, like potential output, is abstract in nature. It is not measured directly but is constructed based on a concept or a definition. Consequently, any measure will depend on how core inflation is defined. Similarly, the optimal measure will depend on the criterion used to assess competing measures of core inflation. In the literature, there are a variety of definitions and criteria used in relation to core inflation.

From the perspective of a central bank, the most useful definition of core inflation is that it represents monetary inflation, which is distinct from headline inflation. Monetary inflation is inflation that is directly influenced by monetary policy. It is conceived as affecting all prices uniformly and represents a common element to all price changes. Headline inflation, as measured the national consumer price index, is generally used as an indicator of changes in the cost of living as its weights are derived on the basis of expenditure shares of a representative basket of goods. The distinction between headline inflation and monetary inflation is made on the basis that monetary inflation determines the price level in the long-run but non-monetary, short-run factors can influence the headline inflation rate in the short-run. The challenge empirically is to distil monetary or core inflation from the headline inflation rate.

Given this definition of core inflation as monetary inflation, its usefulness as a forecasting tool is obvious. The aim of this paper is to estimate measures of core inflation for Ireland and compare their ability to forecast headline inflation against purely statistical alternatives. The first structural VAR measure of core inflation for Ireland is estimated using long-run restrictions. An Autoregressive Integrated Moving Average (ARIMA) model fitted to the headline inflation rate is used to construct the benchmark forecast. The ARIMA forecast is found to be the best way of forecasting headline inflation over very short time horizons. For forecasts over longer horizons, a forecast using a structural VAR measure of core inflation out-performs statistical measures of core inflation put in the same forecasting model. It also out-performs forecasts from the ARIMA benchmark. The next section contains a literature review. Section 3 outlines the structural VAR methodology while section 4 descibes the data used in the study. Section 5 describes the properties of the different core estimators, including the construction the trimmed means. Section 6 details the forecasting criterion used to assess the estimators, which includes the results. Section 7 concludes the paper.

1.2 Literature Review

There are two basic approaches to measuring core inflation. Hogan et al (2001) label one the statistical approach and the other the modelling approach. The statistical approach is a practical, data-driven approach. The problem is to find a measure of core inflation from the data on price indices and inflation rates. The most simple of these approaches is to exclude some component of the consumer price index that is the most volatile. For instance, a common euro area measure of core inflation is the Harmonised Index of Consumer Prices (HICP) excluding energy. In essence, this represents a re-weighting of the HICP with the energy component given a zero weighting. However, energy may not be the most volatile component in every period. Despite this drawback, the HICP excluding energy is included as one of the measures of core inflation in this paper because it is very widely reported and because there is no computational cost.

Macklem (2001) suggests a measure of core inflation that excludes the eight most volatile components of the CPI (out of a total of fifty-four) on the basis of measured average volatility over a number of preceding time periods. This approach is also open to the criticism that the most volatile components in the current period may not be excluded. A more dynamic method is to measure the volatility of all components in each current period and then exclude a certain number. A problem with these approaches is that the excluded items, although volatile, may contain information regarding the core inflation signal. Dow (1994) re-weights the CPI so that the weight of each component is inversely proportional to its variance. In this way, no component with potentially valuable information regarding core inflation is totally excluded. Blinder (1997) also suggests an inclusive measure in which each component is weighted according to its ability to forecast future inflation.

It is also possible to apply a simple statistical smoothing or filtering technique to arrive at a measure of core inflation. A statistical filter generally works on the premise that the inflation rate being examined contains both a trend and a cyclical component. The aim is to "filter" out the cyclical component, leaving only the underlying trend in inflation. Basic techniques, such as standard or centred moving averages, can also be used. The statistical filter used in this study is the Hodrick-Prescott (HP) filter. The main advantage of using a HP filter is that it is well understood in the profession. However, the end-point problem with the HP filter will hinder forecasts to a certain extent.

Another strand of literature in the statistical approach considers the distribution of individual price changes that constitutes the CPI. The key insight in this approach is that the observed price changes are a sample drawn from an unobserved population distribution of price changes. The aim is to estimate the population mean from the observed sample. If the population is normally distributed, the sample mean will be an unbiased and efficient estimator. However, if the population distribution exhibits excess kurtosis, the sample will contain more extreme values than a normal distribution. In this case, the sample mean will not be an efficient estimator of the population mean. In general, as the kurtosis of the distribution increases, the efficiency of estimators - like the sample mean - that place a high weight on observations in the tails of the distribution decreases relative to estimators that place a low weight on the tails of the distribution (Roger, 1998).

In many countries, it has been found that the distribution of price changes is positively skewed with excess kurtosis. Meyler (1999) demonstrates that this characterisation also holds for Irish price changes. Robust or limited-influence estimators have been proposed as the optimal measure of population central tendency in this case. These estimators ignore a certain proportion of the tails of the distribution. Consequently, they aren't influenced by extreme observations. For example, a 10% trimmed mean ignores 5% of the observations at each end of the distribution and takes the mean of the remaining observations. Trimmed means are the most common limited influence estimator but trimmed medians can also be used. Updating the work of Meyler (1999), trimmed means with various levels of trim are estimated in this paper although a slightly different methodology is employed.

The optimal trim depends on the benchmark used. A desirable characteristic of core inflation is that it should track trend inflation. Cecchetti (1997), Kearns (1998) and Meyler (1999) compare their estimates of core inflation to a centred moving average of headline inflation, which is assumed to mimic trend inflation. Another common benchmark is to compare the error from a forecasting model using core inflation against the same forecasts made using headline inflation. Meyler (1999) and Clark (2001) compare forecast errors from an ARIMA model and a simple regression respectively. Forecasting ability is the benchmark used when assessing the optimal level of trim in this paper.

Statistical approaches are often criticised on the grounds that they don't rely on any economic theory. In contrast, structural models of core inflation are heavily grounded in theory. Quah and Vahey (1995) propose a measure of core inflation based on the concept of a vertical Philips curve. Inflation is assumed to be affected by two different types of shock, distinguished by their effect on output. The core inflation shock is output neutral after some fixed horizon whereas the non-core shock is allowed to influence output in the long-run. Core inflation is defined by Quah and Vahey as "the underlying movement in measured inflation associated only with the first kind of disturbance". The methodology has been widely implemented to measure core inflation internationally but has yet to be applied in Ireland.

1.3 Methodology

The study compares four measures of core inflation. The HICP excluding energy measure of core inflation is calculated by the Central Statistics Office. The HP filter is well-understood and needs no explanation. The methodology used in the construction of the trimmed mean measures is detailed in section 5.3.2. This section explains the methodology for the SVAR estimate. The methodology is identical to that used by Quah and Vahey (1995), using the type of long-run restrictions first proposed by Blanchard and Quah (1989) although the exposition of the model generally mirrors that of Claus (1997). The model is formulated in terms of the first differences of oil prices, output and the inflation rate. This has the implication that prices are $I(2)$. In the moving average representation, the series can be expressed as a function of past and present structural shocks:

$$\Delta oil_t = \sum_{k=0}^{\infty} s_{11,k}\, \epsilon_{1t-k} + \sum_{k=0}^{\infty} s_{12,k}\, \epsilon_{2t-k} + \sum_{k=0}^{\infty} s_{13,k}\, \epsilon_{3t-k} \tag{1.3.1}$$

$$\Delta y_t = \sum_{k=0}^{\infty} s_{21,k}\, \epsilon_{1t-k} + \sum_{k=0}^{\infty} s_{22,k}\, \epsilon_{2t-k} + \sum_{k=0}^{\infty} s_{13,k}\, \epsilon_{3t-k} \tag{1.3.2}$$

$$\Delta \pi_t = \sum_{k=0}^{\infty} s_{31,k}\, \epsilon_{1t-k} + \sum_{k=0}^{\infty} s_{32,k}\, \epsilon_{2t-k} + \sum_{k=0}^{\infty} s_{33,k}\, \epsilon_{3t-k} \tag{1.3.3}$$

where oil_t, y_t and π_t denote the logs of oil prices, output and the inflation rate respectively. The three structural shocks ϵ_{1t}, ϵ_{2t} and ϵ_{3t} can be thought of as an oil price shock, a non-core shock and a core shock respectively. These shocks are orthogonal, white noise errors. This type of model is frequently modelled with a bivariate specification using only output and inflation but the openness of the Irish economy suggests some role for external shocks in the system. For this reason, oil prices were also chosen from a selection of open economy variables. An alternative would be to choose the policy rate as the third variable and then define core inflation as that part of inflation not affected by interest rate shocks but the policy rate has remained unchanged in a large number of months in the sample. In matrix form, this system can be written:

$$\begin{bmatrix} \Delta oil_t \\ \Delta y_t \\ \Delta \pi_t \end{bmatrix} = \begin{bmatrix} S_{11}(L) & S_{12}(L) & S_{13}(L) \\ S_{21}(L) & S_{22}(L) & S_{23}(L) \\ S_{31}(L) & S_{32}(L) & S_{33}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} \tag{1.3.4}$$

or

$$X_t = S(L)\epsilon_t \tag{1.3.5}$$

where $S(L)$ is a polynomial in the lag operator whose individual coefficients are denoted $s_{ij,k}$. The structural shocks are normalized so that their covariance matrix is the identity matrix. It is the behaviour of the structural shocks, which represent the core and non-core inflation shocks, that is really of interest. The problem is that, in the estimation of a standard reduced-form VAR, it is the reduced-form shocks and not the structural shocks that are estimated. Nonetheless, the first step in identifying the structural shocks is the estimation of the reduced-form VAR. Ignoring the intercept for simplicity:

$$\begin{bmatrix} \Delta oil_t \\ \Delta y_t \\ \Delta \pi_t \end{bmatrix} = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta_{21} & \Theta_{22} & \Theta_{23} \\ \Theta_{31} & \Theta_{32} & \Theta_{33} \end{bmatrix} \begin{bmatrix} \Delta oil_{t-1} \\ \Delta y_{t-1} \\ \Delta \pi_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} \tag{1.3.6}$$

or

$$X_t = \Theta X_{t-1} + e_t \tag{1.3.7}$$

Assuming that Θ is invertible, the Wold moving average representation can be obtained:

$$\begin{bmatrix} \Delta oil_t \\ \Delta y_t \\ \Delta \pi_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \\ C_{31}(L) & C_{32}(L) & C_{33}(L) \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} \tag{1.3.8}$$

or

$$X_t = C(L)e_t \tag{1.3.9}$$

where $C(L)$ is a polynomial in the lag operator. This means X_t can be expressed:

$$X_t = e_t + \Theta e_{t-1} + \Theta^2 e_{t-2} + ... \tag{1.3.10}$$

The matrix $C(1)$ is the matrix of long-run effects with respect to the reduced-form shocks.

$$\begin{aligned} C(1) &= \sum_{k=0}^{\infty} C_k L^k, \;\; C_0 = I_3, C_k = \Theta^k && (1.3.11) \\ &= (I_3 - \Theta L)^{-1} && (1.3.12) \end{aligned}$$

The reduced-form shocks are a linear combination of the structural shocks:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} = \begin{bmatrix} s_{11}(0) & s_{12}(0) & s_{13}(0) \\ s_{21}(0) & s_{22}(0) & s_{23}(0) \\ s_{31}(0) & s_{32}(0) & s_{33}(0) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} \tag{1.3.13}$$

or

$$e_t = S(0)\epsilon_t \tag{1.3.14}$$

Given this relationship between the structural and reduced-form shocks, equation (13) can be re-written in terms of the structural shocks as follows:

$$X_t = S(0)\epsilon_t + \Theta S(0)\epsilon_{t-1} + \Theta^2 S(0)\epsilon_{t-2} + ... \tag{1.3.15}$$

The elements of the matrix $S(0)$ are still unknown. The matrix contains nine elements. Thus, nine independent equations are needed in the nine elements. Consider the variance/covariance matrix of the reduced-form residuals:

$$\Sigma = E(e_t e_t{}') = S(0)E(\epsilon_t \epsilon_t{}')S'(0) = S(0)S'(0) \tag{1.3.16}$$

The values of Σ are known from the estimation of the reduced-form VAR. This allows us to write six equations in terms of the nine unknowns:

$$\begin{aligned} var\,(e_{1t}) &= s_{11}(0)^2 + s_{12}(0)^2 + s_{13}(0)^2 && (1.3.17) \\ var\,(e_{2t}) &= s_{21}(0)^2 + s_{22}(0)^2 + s_{23}(0)^2 && (1.3.18) \\ var\,(e_{3t}) &= s_{31}(0)^2 + s_{32}(0)^2 + s_{33}(0)^2 && (1.3.19) \\ cov\,(e_{1t},\; e_{2t}) &= s_{11}(0)s_{21}(0) + s_{12}(0)s_{22}(0) + s_{13}(0)s_{23}(0) && (1.3.20) \\ cov\,(e_{1t},\; e_{3t}) &= s_{11}(0)s_{31}(0) + s_{12}(0)s_{32}(0) + s_{13}(0)s_{33}(0) && (1.3.21) \\ cov\,(e_{2t},\; e_{3t}) &= s_{21}(0)s_{31}(0) + s_{22}(0)s_{32}(0) + s_{23}(0)s_{33}(0) && (1.3.22) \end{aligned}$$

In order to get the remaining equations, explicit restrictions are placed on the long-run behaviour of the system. The long-run effects of the reduced form shocks were given by the matrix $C(1)$. Equation (14) gives the relationship between the reduced form shocks and the structural shocks. This allows the long-run effects of the structural shocks, denoted by the matrix $S(1)$, to be expressed as follows:

$$\begin{bmatrix} S_{11}(1) & S_{12}(1) & S_{13}(1) \\ S_{21}(1) & S_{22}(1) & S_{23}(1) \\ S_{31}(1) & S_{32}(1) & S_{33}(1) \end{bmatrix} = \begin{bmatrix} C_{11}(1) & C_{12}(1) & C_{13}(1) \\ C_{21}(1) & C_{22}(1) & C_{23}(1) \\ C_{31}(1) & C_{32}(1) & C_{33}(1) \end{bmatrix} \begin{bmatrix} s_{11}(0) & s_{12}(0) & s_{13}(0) \\ s_{21}(0) & s_{22}(0) & s_{23}(0) \\ s_{31}(0) & s_{32}(0) & s_{33}(0) \end{bmatrix} \quad (1.3.23)$$

or

$$S(1) = C(1)S(0) \quad (1.3.24)$$

If the matrix $S(1)$ is lower triangular, the necessary equations for identification can be found from the resulting restrictions. These restrictions impose structure on the economic relationships between the variables in the system. The first restriction is that $S_{23}(1) = 0$ and this amounts to saying that the core shock has no effect on output in the long-run. This is consistent with the idea of a vertical long-run Philips curve and is a traditional identifying assumption in the application of long-run restrictions. The next two restrictions are that $S_{12}(1)$ and $S_{13}(1) = 0$. The implication of these restrictions is that domestic core and non-core shocks have no influence on international oil prices in the long-run. Bjornland (2001) justifies the use of these restrictions in the case of Norway on the basis that it is a small oil producer with limited influence on oil prices. The same restrictions for Ireland are even less contentious given that we are a small oil-importing economy. These three restrictions yield the following equations:

$$C_{11}(1)s_{12}(0) + C_{12}(1)s_{22}(0) + C_{13}(1)s_{23}(0) = 0 \quad (1.3.25)$$

$$C_{11}(1)s_{13}(0) + C_{12}(1)s_{23}(0) + C_{13}(1)s_{33}(0) = 0 \quad (1.3.26)$$

$$C_{21}(1)s_{13}(0) + C_{22}(1)s_{23}(0) + C_{23}(1)s_{33}(0) = 0 \quad (1.3.27)$$

It is now possible to estimate all elements of $S(0)$. Together with $C(1)$, which is calculated from the reduced-form coefficients, this allows the structural shocks to be identified.

1.4 Data

Both monthly and quarterly data are used to calculate a SVAR measure of core inflation in the paper. The inflation rate considered is the year-on-year change in the Harmonised Index of Consumer Prices (HICP). Output is measured using the seasonally adjusted industrial production index for monthly data and an interpolated measure of real GDP for quarterly data. Oil prices refer to the price of UK Brent. The monthly data are available over the period 1997M1-2006M5. This is a relatively short sample in the context of a SVAR model imposing long-run restrictions but the results from the model appear reasonable. Despite the short sample, it is the results of the monthly analysis that are of most interest because future trends in inflation are most likely to be spotted first from monthly data rather than quarterly data. The inclusion of quarterly data allows the evolution of core inflation to be tracked over a longer period. The monthly data relate to a period when the economy has been in a state of perpetual boom. However, the quarterly data set spans 1980Q1-2006Q2 so it also contains data on a period when the economy was underperforming. The macro series are examined in greater detail in the appendix.

In terms of constructing a trimmed mean, the process is data-intensive. The monthly SVAR data span two inflation base periods. The first base period covers the years 1997-2001 while the second base period covers 2002-present.[1] In the first base period, the HICP has 529 individual price series. This increases to 606 individual series for the second based period. This is a much wider cross section of data that has been available in other comparable studies. The change in the number of individual price series is not solely due to additional items being included in the representative basket of consumer goods; items are also replaced and deleted.

1.5 Overview of Core Inflation Measures

1.5.1 HICP excluding Energy

The first measure of core inflation considered is the HICP excluding energy. This measure of core inflation will only differ from the headline rate to a meaningful degree when there are large changes in energy prices. Figures 1 and 2 graph this measure of core inflation for both monthly and quarterly data. There are few instances of a large sustained divergence between the two series although the effect of high energy prices in the past two years is quite noticeable, particularly from the monthly data. To the extent that the core series is so similar to the HICP, it might not be expected to provide much additional informational content for forecasting headline inflation that is not contained in the headline rate itself.

1.5.2 Hodrick-Prescott Filter

The Hodrick-Prescott filter is used as the second measure of core inflation. The value of the smoothing parameter, λ, is chosen in order to minimise the errors from a forecasting regression, which is presented later. Figures 3 and 4 graph the headline inflation rate and the HP filtered measure of core inflation for both monthly and quarterly data. The HP measure of core inflation tracks the headline inflation rate in a much smoother fashion than the HICP excluding energy. The difference between the two series alternates from positive to negative quite frequently. The filter is purely mechanical however. It attributes a certain proportion of each shock hitting the series to a change in the trend of the series while the remainder is regarded as temporary noise. As with the HICP excluding energy, there is no structural interpretation to this core measure.

[1] The present base period will run until the end of 2006.

1.5.3 Trimmed Means

Properties of Price Change Distributions

It was mentioned that the key motivation for the construction of trimmed mean estimates of core inflation is that the sample mean is an inefficient estimator of population central tendency when the sample exhibits excess kurtosis. Table 1 provides a summary of some of the key properties of the sample distribution of price changes. The trimmed means are estimated for the span of the monthly data only. Results are presented for both month-on-month and year-on-year price changes although the year-on-year statistics are of more interest because the year-on-year inflation rate is included in the SVARs. The summary statistics are calculated for both base periods individually and for the sample as a whole. The change from one base period to another presents difficulties when dealing with the year-on-year price changes. At the start of the second base period, new items are introduced, old items are deleted and other items are replaced. This means that some items do not have a comparator from twelve months earlier from which to calculate a year-on-year change. (This problem does not exist with the month-on-month changes because there is a one month overlap in base periods.) Thus the full sample statistics for the year-on-year price changes include a one year gap. When the trimmed means are calculated, year-on-year approximations are estimated from the monthly data for the one year gap.

The statistics in Table 1 are all averages. The mean, median, skew and kurtosis of the price change distribution are calculated each month in the sample and the results presented are sample averages. On examination of national price change data, numerous researchers have found price change distributions to be characterised by positive skew. Table 1 indicates that the month-on-month price change distributions are also characterised by positive skew for Ireland. The year-on-year price change distribution for the full sample is broadly symmetric with a small negative skew in the first base period largely offset by a similar positive skew in the second base period. The larger skew in month-on-month inflation rates may point to price skickiness in the short-run with the absence of skew in year-on-year rates showing that prices are flexible downwards over longer horizons.

Excess kurtosis is an obvious feature of all distributions. It is more pronounced in the case of month-on-month price changes but it is still a significant feature of the data in the year-on-year case. The kurtosis of the distribution is more readily apparent from graphical evidence. As an example, Figure 5 graphs the year-on-year price change distribution for January 2003 overlaid with a normal density using the sample mean and variance. A distribution with excess kurtosis relative to the normal distribution has a more acute peak around the mean and more weight in the tails. The peak in Figure 5 is clearly higher than the normal distribution. The mean price change is 1.5% with a standard deviation of 7.2%. A 99% confidence interval for a normal distribution with these moments is approximately -17% to 20%. However, it is clear from the graph that more than 1% of the distribution lies outside this interval, which is further evidence of the kurtosis of the distribution. The median is 1.9%, slightly higher than mean, resulting in a small negative skew. Figure 6 presents a similar graph for November 2005. It indicates that excess kurtosis is also a feature of the data for months characterised by positive skew. The kurtosis of these distributions warrant the use of trimmed means as measures of core inflation.

Constructing the Trimmed Mean Measures

The trimmed mean can be calculated in two different ways. The most common approach is to estimate the inflation rates of all the individual components that comprise the HICP and then rank these inflation rates and their associated expenditure weights. Exclude the items associated with a certain percentage of the largest and smallest inflation rates. Calculate the aggregate inflation rate of the remaining items, rescaling the weights used to calculate the headline inflation rate so that the new weights still sum to 1. Studies of core inflation that report a trimmed mean often report the result of this sort of calculation.

The problem with this sort of approach is that the weights are based on expenditure shares of a representative basket of goods, devised by statistical agencies to approximate changes in the cost of living. There is no reason to believe that this weighting system should still be used when constructing a core inflation measure, which aims to capture the underlying trend in inflation rather than the cost of living. In fact, the weighting system will have a large distortionary effect on the underlying inflation signal if price changes due to idiosyncratic shocks occur in items with large expenditure weights. Thus, a second method to calculate trimmed means simply ignores the weights and calculates a simple average of individual inflation rates following the trimming operation. As before, begin by ordering individual inflation rates and excluding a certain percentage but, this time, take a simple average of the remaining inflation rates. This could be referred to as a simple trim to distinguish it from the standard trimming method and it is the method employed in this paper.

Figure 7 plots trimmed means with 5% and 10% trims. Both trimmed mean measures of core inflation are substantially lower than the headline rate of inflation for most of the sample. On average, the 5% trimmed mean is 1.9% lower than headline inflation while the 10% trimmed mean is 1.8% lower. Figure 9 plots the average inflation rate without any trim and the median inflation rate. These two series broadly resemble the trimmed mean series. Average inflation is consistently lower than headline inflation. This indicates that the weighting system used to calculate headline inflation has contributed to the relatively high rate of inflation over much of the sample.

1.5.4 Structural VAR Estimates

The monthly data span the period 1998M1 - 2006M5. The structural VAR is formulated using a trivariate specification with oil prices, industrial production and the inflation rate. The variables enter the model in first difference form. The results of the unit root tests for the monthly data are presented in Table 2. The unit root test performed is the Elliot-Rothenberg-Stock (1996) DF GLS test with the lag length for the test chosen on the basis of the modified AIC suggested by Ng and Perron (2001). The first difference of inflation and oil prices are found to be stationary. However, on the basis of this test, the first difference of industrial production is non-stationary. Measures of economic output were very high in the early part of the sample due to the extraordinary economic growth at that time. The average value of the growth rate of industrial production for 1998-2001 is 13.9% while the corresponding figure for 2002-2006M5 is just 3.8%. The difference between these two figures suggests that there may have been a structural break in the series.

The Sup F test is performed to test for a structural break in GLS ADF detrending regression for industrial production growth. The null hypothesis of structural stability is rejected at the 1% level. The least squares estimator of the break date is the date which maximises the Chow test statistic. This estimates the break in the series at 2001M4. In the context of this break, Perron's (1989) unit root test in the presence of a structural break is carried out in preference to the DF GLS test. The null hypothesis that the first difference of the industrial production series has a unit root is rejected. Thus, all series enter the VAR in first difference format. Table 3 presents the results of cointegration tests for the three series. The results are based on a test for one cointegrating vector amongst the three variables and indicate that the series are not cointegrated. Inflation and output were also tested for a cointegrating vector separately. Again, no evidence of cointegration was found but results are not presented for the sake of brevity.

The maximum lag length considered for the VAR is the frequency of the data plus one. The VAR is specified with four lags. The number of lags was chosen to maximise the forecasting ability of the resulting core measure. The core inflation measure is not sensitive to small changes in the number of lags specified in the VAR. Given the presence of a structural break in the output series, the VAR was originally specified with a dummy variable to capture this effect. However, the short length of the sample combined with the recursive process that is used to perform forecasts meant that the break parameter was poorly estimated in most samples and this resulted in larger forecast errors. For this reason, the final specification of the SVAR excludes the structural dummy variable. Figure 9 graphs the SVAR measure of core inflation using monthly data. The graph shows that this measure of core inflation largely tracks the headline inflation rate for most of the sample.

The quarterly SVAR also uses a trivariate specification but GDP is used as the output variable rather than industrial production. The results of the unit root tests for the quarterly variables are also presented in Table 2. Energy prices are again found to be I(1) but the year-on-year inflation rate calculated using quarterly data is found to be I(0) despite a high rate of inflation in the early eighties. This means that the inflation rate enters the VAR in levels rather than in first differences. GDP growth is found to I(1) but, as is the case with the monthly data, the series is again found to have a structural break. The break date is estimated at 1995Q1. From 1980Q1 up to 1995Q1, economic growth was low, particularly in the eighties, whereas growth over the last eleven years has been very robust. Accounting for this structural break, Perron's test still fails to reject the unit null at the 5% level but comes close to rejection at the 10% level. The power to reject the null is reduced by the small size of the sample however and it appears that the series is more appropriately described as stationary so the series is entered in the VAR in first difference form despite the test results. Again, Table 3 indicates that the three variables are not cointegrated.

Figure 10 graphs headline inflation and the quarterly SVAR measure of core inflation. In the early part of the sample, the two series are broadly similar. However, core inflation is either higher than or broadly similar to headline inflation in the period from 1995-2002. This reflects the fact that economic growth was exceptionally high over this period. Consider a measure of economic growth calculated as the average of the year-on-year growth rate in real GDP for the four most recent quarters. The average value of this growth rate was just over 9 percent between the first quarter of 1995 and the last quarter of 2002. The average growth rate for the remainder of the sample is roughly half that at 4.5 per cent. Core inflation could be expected to be high during this high growth period in the sample. The core inflation series is lower than the headline inflation rate over the last few years and this can be attributed to high oil price inflation.

Having now constructed all core inflation rates, the estimators for both monthly and quarterly data are presented together in Figures 11 and 12 in order to examine the similarity and differences between the series. Clearly, there are significant differences between the monthly estimates. The trimmed series are systematically lower than the other core series. The HP filter, the HICP excluding energy and the SVAR are closer in terms of the estimates of the core inflation rate but significant differences still emerge between these series at certain points in the sample. The correlations between the core series range from 0.99 for the two trims to 0.63 for the correlation between the 10% trim and the SVAR. Clearly, the estimators have their own properties so some way is needed to choose between them. As there are no trims for the quarterly data, the differences between the core estimators are less pronounced. The correlation between the series is at least 0.97 in all cases. This might suggest there is little to choose between them but, if we take the absolute difference between the svar and HP filer, the average value of this series is 0.85 over the sample from 1990:1-2006:2. Thus, even omitting the early, variable part of the sample, there is on average a 0.85% difference in the estimate of core inflation, which is quite sizable and again suggests a formal method of evaluating the series is needed. The evaluation criterion used is forecast ability.

1.6 Forecasting Ability of Core Inflation Measures

In this section, competing measures of core inflation are ranked according to their ability to forecast the headline inflation rate. This is accomplished using a simple forecasting regression:

$$\pi_{t+h} - \pi_t = \alpha + \beta \left(\Pi_t - \pi_t \right) + v_t \tag{1.6.28}$$

where π_t is the inflation rate at time t and Π_t is core inflation. The left hand side of the equation is the difference between headline inflation today and headline inflation h periods in the future. On the right hand side, the term in brackets is the difference between core inflation and headline inflation. The basic premise of this forecasting regression is that difference between headline inflation and core inflation today has predictive power for headline inflation tomorrow. In particular, if there is a large divergence between headline inflation and core inflation, you would expect headline inflation to move back towards core inflation because core inflation is a measure of the general trend in inflation. This is very much the standard forecasting equation used in the literature - the majority of studies that test the forecasting ability of core inflation use this equation.

The regression computes a forecast over a fixed horizon. For example, using monthly data and setting $h = 12$ would yield a forecast of headline inflation twelve months in the future but would not forecast inflation in the intervening periods. There are two ways to get a continuous forecast to the end of the forecasting horizon. Estimate twelve regressions of the type above setting $h = 1...12$. Alternatively, using only the coefficients from the twelve step ahead regression, the forecast for $t + 12$ months ahead can be estimated using the difference between headline inflation and core inflation in period t. Next, the forecast for $t + 11$ months ahead can be estimated using the difference between core inflation and headline inflation in period $t - 1$. Proceeding accordingly, a full set of forecasts can be computed. Forecasts have been computed using both methods and the forecasts calculated using the first approach have the smallest forecast errors for all core measures and over virtually all time horizons. Consequently, the duplicate set forecast errors from the other approach is not reported.

The monthly forecasts are performed up to twelve months in the future whereas the quarterly forecasts are performed up to two years in the future. The forecasts are performed on a recursive basis, with one observation added to the sample each time. The first sample for the monthly estimates is 1998M1-2003M6. The core inflation measures are calculated over this sample and forecasts are performed for the twelve months up to 2004M6. The process is repeated adding one observation each time so by the end of the final estimation period of 2005M5, there are 24 sets of forecasts for each estimation method. An analogous process is used with the quarterly data. The first estimation sample spans 1981Q1-1999Q4 and 16 sets of forecasts are calculated by again adding one observation to the sample at each step.

The forecasts are evaluated using the Root Mean Square Error (RMSE) from pseudo out-of-sample forecasts as the loss function. An ARIMA model is fitted to the headline rate and this is used to construct the benchmark forecast. Table 4 presents the RMSE from the different forecasting regressions over a twelve month forecast horizon while Figure 13 plots the same data. The unbroken line in Figure 13 shows the forecast errors from the ARIMA model. Over the first two months, the forecast errors from the ARIMA model are lower than those from the core inflation measures. The ARIMA model provides a good short-term forecast. Beyond six months, however, the ARIMA models result in the largest forecast errors. Poor forecast performance over longer horizons is a typical feature of univariate forecasting. With the exception of the first two months, the SVAR measure of core inflation results in the lowest forecast errors. The HP filter and the HICP excluding energy forecasts perform better than the ARIMA forecast beyond a six month horizon but still not as well as the SVAR. They are relatively few papers that use an ARIMA model rather than an AR as a benchmark forecast. One exception is Smith (2004), which finds that weighted median inflation can outperform an ARIMA benchmark for US inflation but finds that the excluding energy measure does not. Thus, the excluding energy measure does slightly better in this application. It appears that the informational content in the structural model allows it to out-perform the univariate forecast and the purely statistical measures of core inflation over most horizons.

Table 5 presents the RMSE of the quarterly forecasts over the two year forecast horizon and the corresponding series are graphed in Figure 14. Again, the solid line represents the graph from the ARIMA benchmark. In the case of quarterly data, the ARIMA benchmark forecast performs well. It has lower forecast errors than both the HICP excluding energy and the HP filter forecasts over the entire forecast horizon. It also out-performs the SVAR measure over the first half of the forecast horizon. The forecast errors from the SVAR measure are lower than the benchmark for the second half of the forecast horizon but the improvement in forecast accuracy is not as large as it is in the monthly case. The results of the quarterly forecast exercises reinforce the usefulness of the SVAR measure of core inflation in forecasting the headline rate over the longer time horizons.

1.7 Summary and Conclusions

The paper set out to evaluate the ability of different core inflation measures to forecast the headline inflation rate. The four measures included are the HICP excluding energy, the HICP filtered using the Hodrick-Prescott filter, trimmed mean measures of core inflation (which also considered average inflation) and a structural VAR model of core inflation estimated using long-run restrictions. An ARIMA model was used to construct the benchmark forecast. The results from models constructed using both monthly and quarterly data indicate that the SVAR measure of core inflation used in the forecasting regression provide the best forecasts over long horizons. However, the SVAR model is slightly out-performed by the ARIMA forecast over short time horizons, implying a role for the ARIMA models in short-term forecasting.

1.8 Figures and Tables

Table 1: Properties of Price Change Distributions

Period	Mean	Median	Skew	Excess Kurtosis
Month-on-Month				
1997-2002	0.00205	0.00163	1.21013	49.43845
2002-2005	0.00076	0.00071	1.10814	27.71529
1997-2005	0.00148	0.00122	1.11435	31.91969
Year-on-Year				
1998-2002	0.02024	0.02469	-0.44942	7.87065
2003-2005	0.00293	0.00409	0.47883	5.60401
1998-2005	0.01294	0.01600	-0.05799	9.59957

Table 2: **Unit Root Tests**

Variable	**Frequency**	**Test Type**	**Setup**$^{+}$	**Statistic**	**5 Percent Critical Value**
π_t	Monthly^{++}	DF GLS	c,2	-0.96	-1.95
$\Delta\pi_t$	Monthly	DF GLS	c,9	-2.00	-1.95
Y_t	Monthly	DF GLS	c,13	-0.88	-1.95
ΔY_t	Monthly	DF GLS	n,13	-1.69	-1.95
$\$ \, oil$	Monthly	DF GLS	c,1	-0.04	-1.95
$\Delta \, \$ \, oil$	Monthly	DF GLS	n,11	-2.18	-1.95
$log(HICP)$	Quarterly^{+++}	DF GLS	t,5	-2.10	-2.89
π_t	Quarterly	DF GLS	n,1	-3.11	-1.95
Y_t	Quarterly	DF GLS	t,5	-1.17	-2.89
ΔY_t	Quarterly	DF GLS	c,5	-1.28	-1.95
$\$ \, oil$	Quarterly	DF GLS	c,5	-0.75	-1.95
$\Delta \, \$ \, oil$	Quarterly	DF GLS	c,2	-5.09	-1.95
ΔY_t	Monthly	Perron	na	4.94	-3.76
Y_t	Quarterly	Perron	na	3.15	-3.76

+ letter refers to detrending used: c = constant, t = constant and trend,
n = none; integer = number of lags used in unit root test
++ monthly sample period: 1998(2) - 2006(5)
+++ quarterly sample period 1980(2) - 2006(2)
Lag length determined using modified AIC.

Table 3: Engle-Granger Cointegration Test

Frequency	**Statistic**	**Critical Value**	**Decision**
Monthly^{+}	-3.42	-3.83	Not Cointegrated
Quarterly^{++}	-1.44	-3.83	Not Cointegrated

+ Cointegrating vector: $(Y_t, \pi_t, \$ \, oil_t)$
++ Cointegrating vector: $(Y_t, \log(HICP_t), \$ \, oil_t)$
Same sample lengths as previous table.

Table 4: RMSE from Monthly Inflation Forecasts

Forecast Horizon	Forecast Method				
Months	ARIMA	SVAR	Exc. Energy	HP Filter	Median
1	0.26604	0.28655	0.29661	0.30256	0.28987
2	0.44016	0.46988	0.50140	0.53139	0.48863
3	0.60978	0.59405	0.65738	0.64045	0.63319
4	0.79296	0.73949	0.82787	0.78211	0.81187
5	0.94063	0.85301	0.95916	0.91707	0.95926
6	1.08941	0.94338	1.06291	1.02833	1.09170
7	1.20568	0.99713	1.15083	1.09903	1.18524
8	1.30520	0.99130	1.16362	1.10268	1.21926
9	1.36899	0.96241	1.14462	1.06885	1.24982
10	1.41873	0.88954	1.08209	0.99439	1.26472
11	1.48252	0.82169	1.05068	0.95102	1.31448
12	1.58602	0.85291	1.10732	1.02767	1.45748

Table 5: RMSE from Quarterly Inflation Forecasts

Forecast Horizon	Forecast Method			
Quarters	ARIMA	SVAR	Exc. Energy	HP Filter
1	0.65709	0.76325	0.86282	1.09124
2	0.84520	0.95420	1.23127	1.62088
3	0.92243	1.03994	1.24845	1.83381
4	1.15184	1.22585	1.37805	2.6128
5	1.21559	1.22306	1.35116	2.30343
6	1.22766	1.14692	1.24931	2.09431
7	1.26532	1.17150	1.37351	1.95942
8	1.30851	1.21585	1.48887	1.61224

Figure 1: Inflation and Inflation excluding Energy

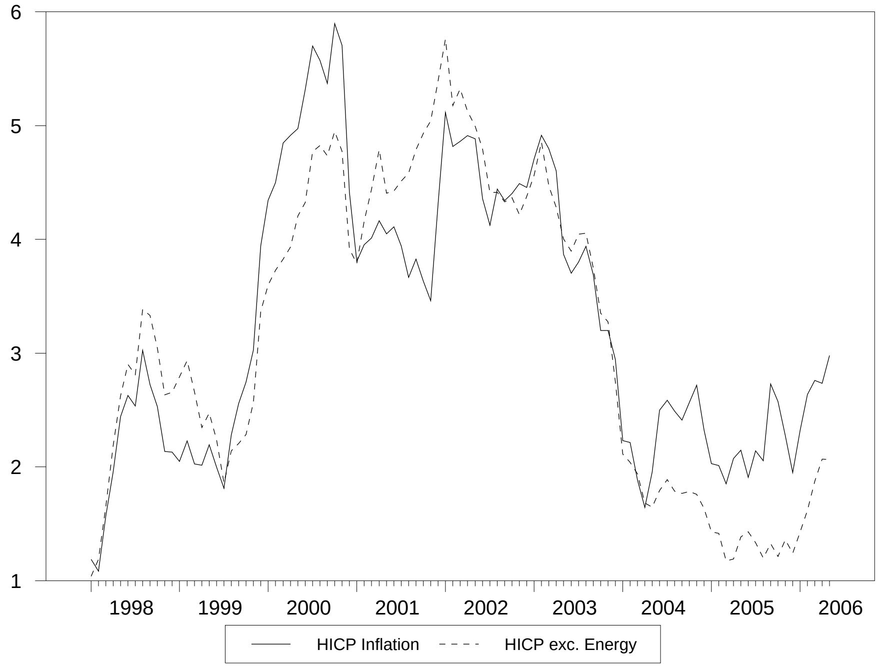


Figure 2: Inflation and Inflation excluding Energy

Quarterly Data

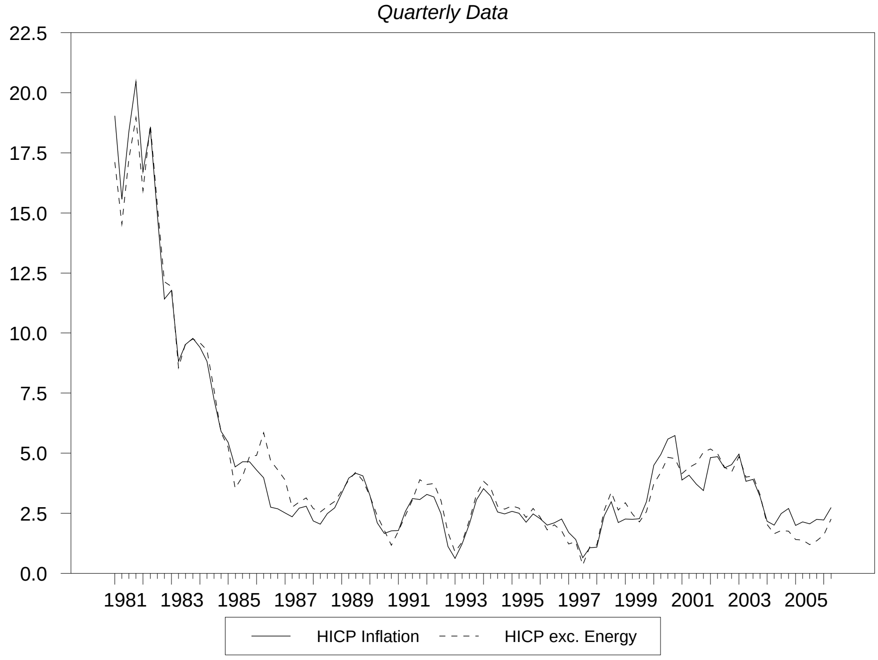


Figure 3: Headline Inflation and HP Filter

Monthly Data

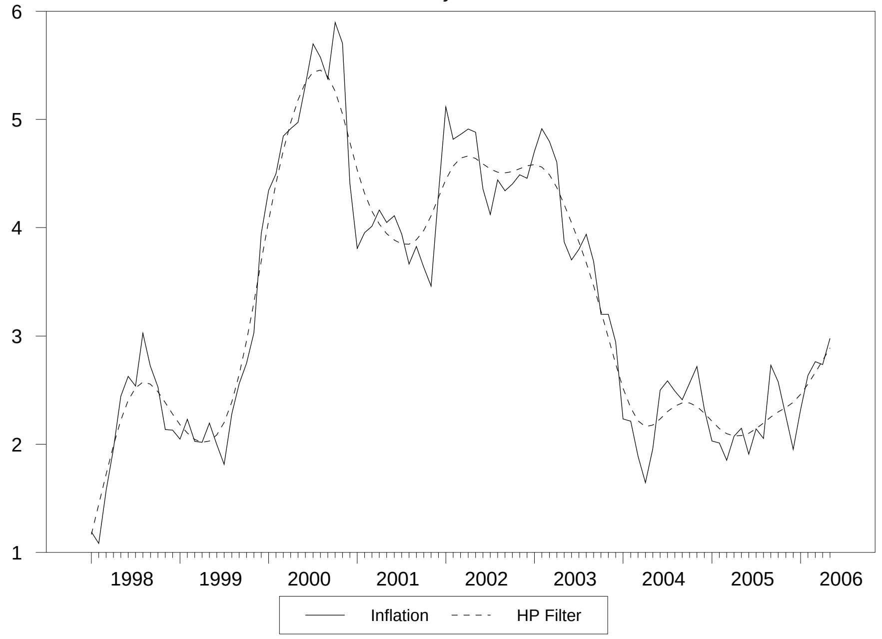


Figure 4: Headline Inflation and HP Filter

Quarterly Data

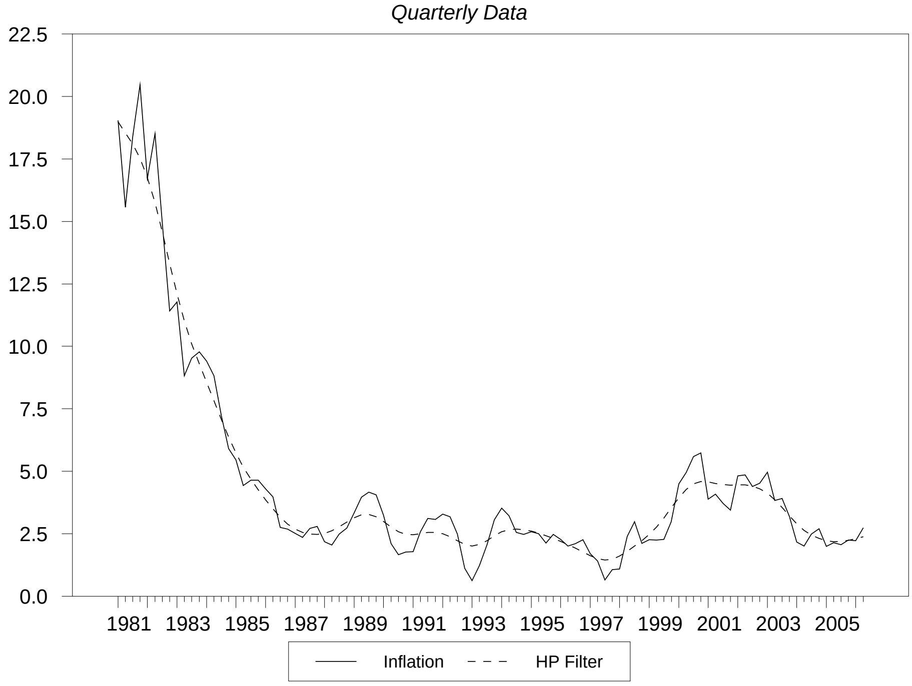


Figure 5: Distrubution of Year-on-Year Price Changes

Period: January 2003

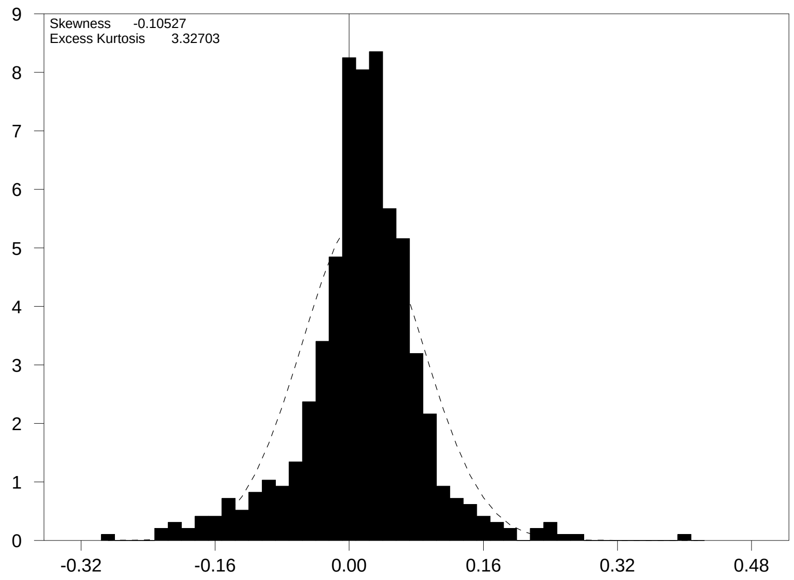


Figure 6: Distrubution of Year-on-Year Price Changes

Period: November 2005

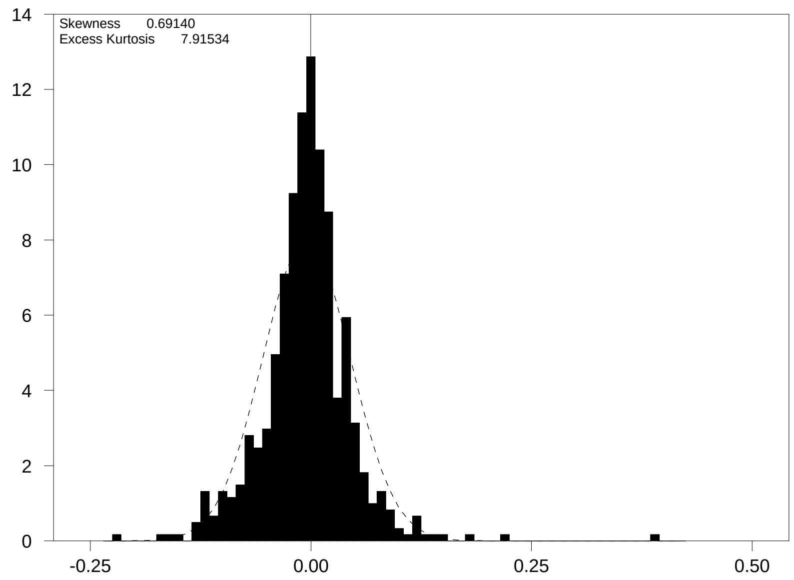


Figure 7: Headline Inflation, 5% and 10% Simple Trim

Monthly Data

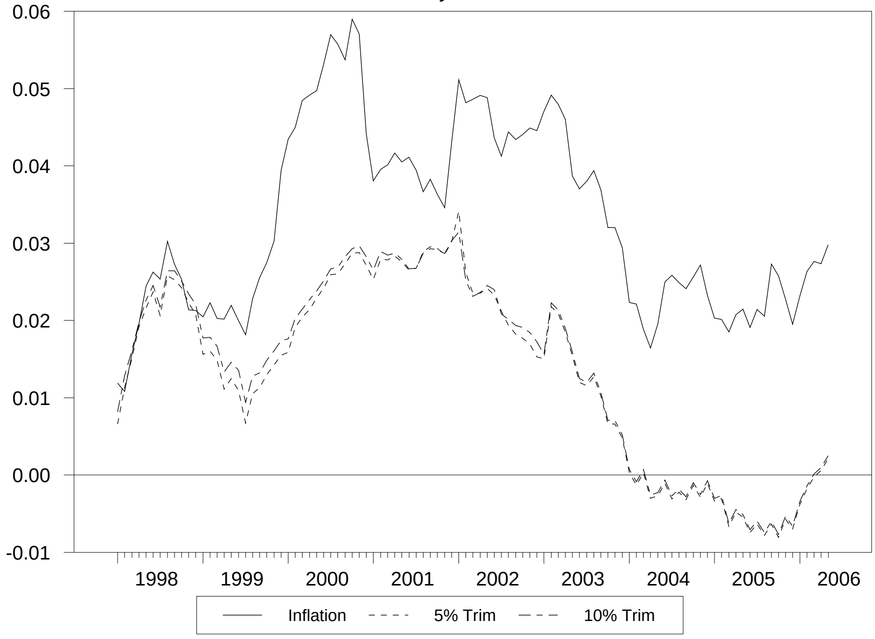


Figure 8: Headline, Median and Average Inflation Rates

Monthly Data

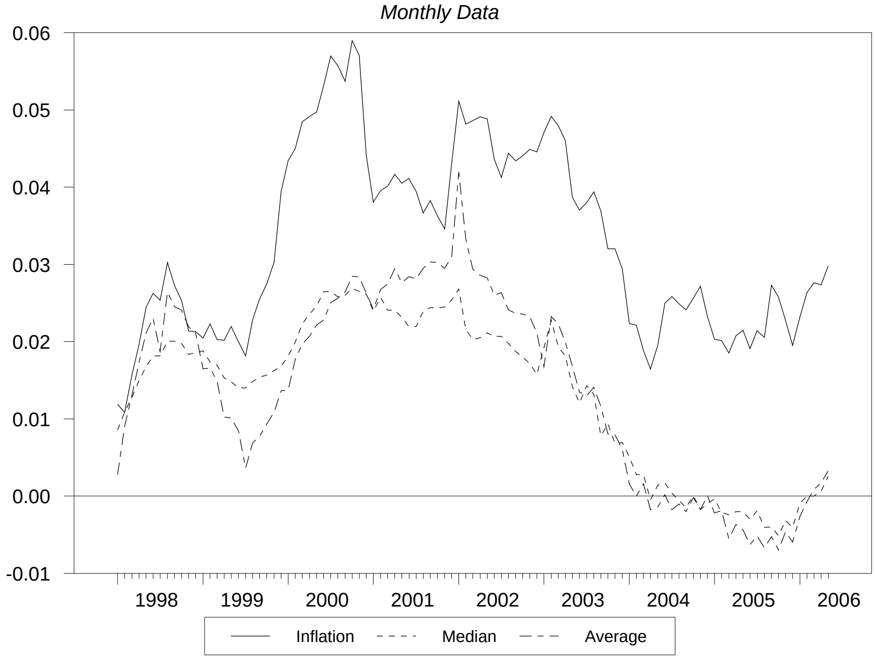


Figure 9: Headline Inflation and SVAR Core Inflation

Monthly Data

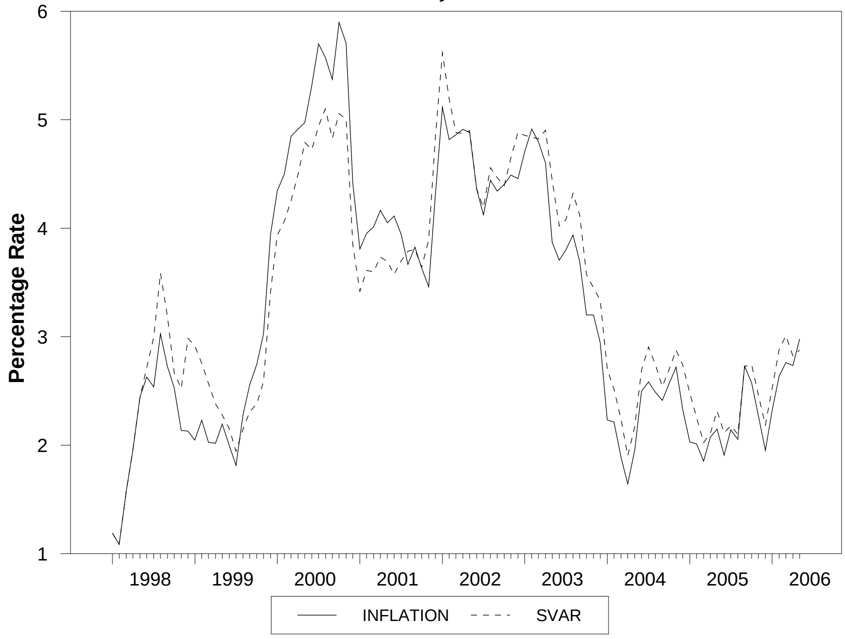
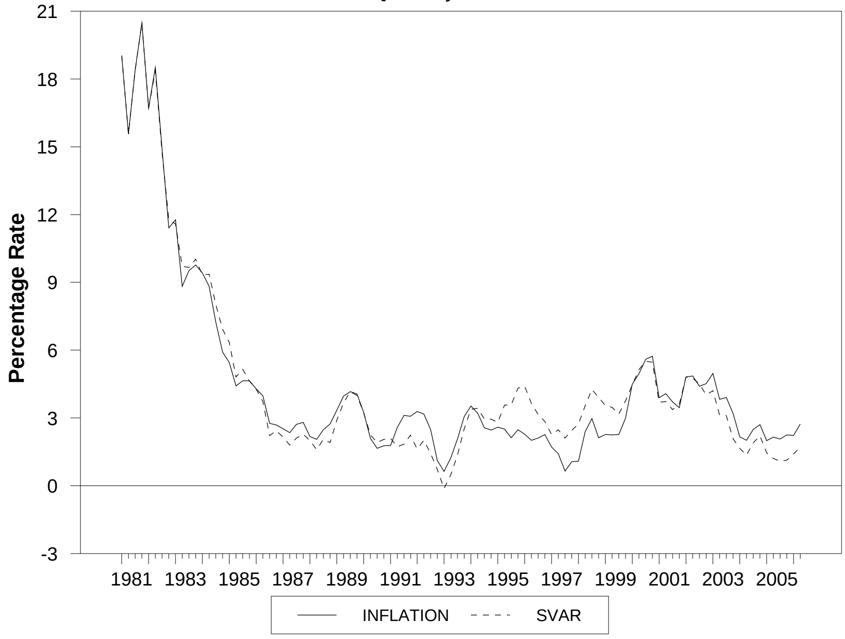


Figure 10: Headline Inflation and SVAR Core Inflation

Quarterly Data



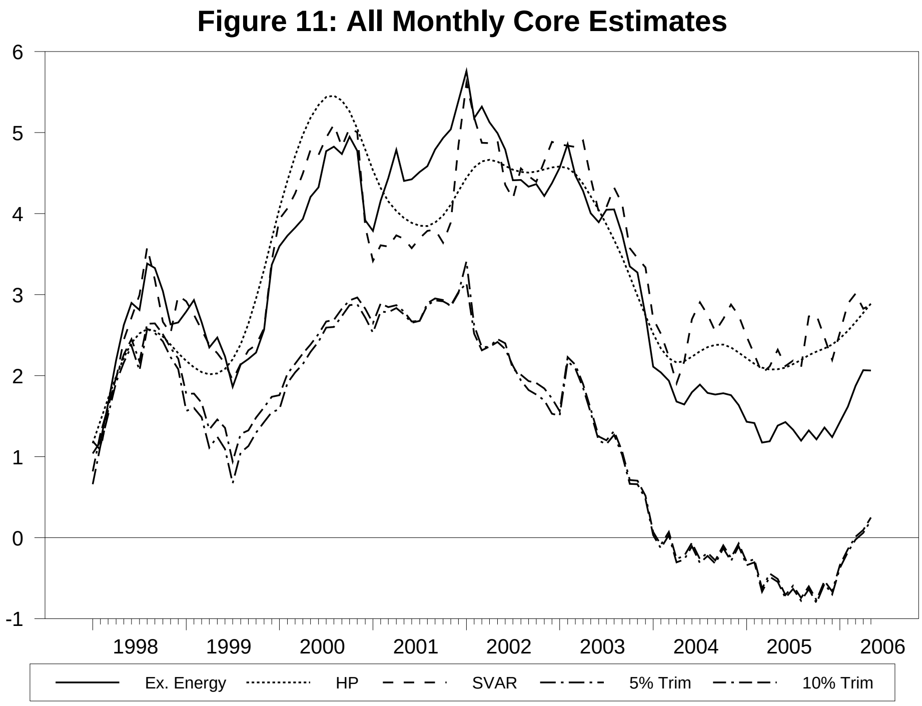


Figure 11: All Monthly Core Estimates
6
5
4
3
2
1
0
-1
1998
1999
2000
2001
2002
2003
2004
2005
2006
Ex. Energy
HP
SVAR
5% Trim
10% Trim

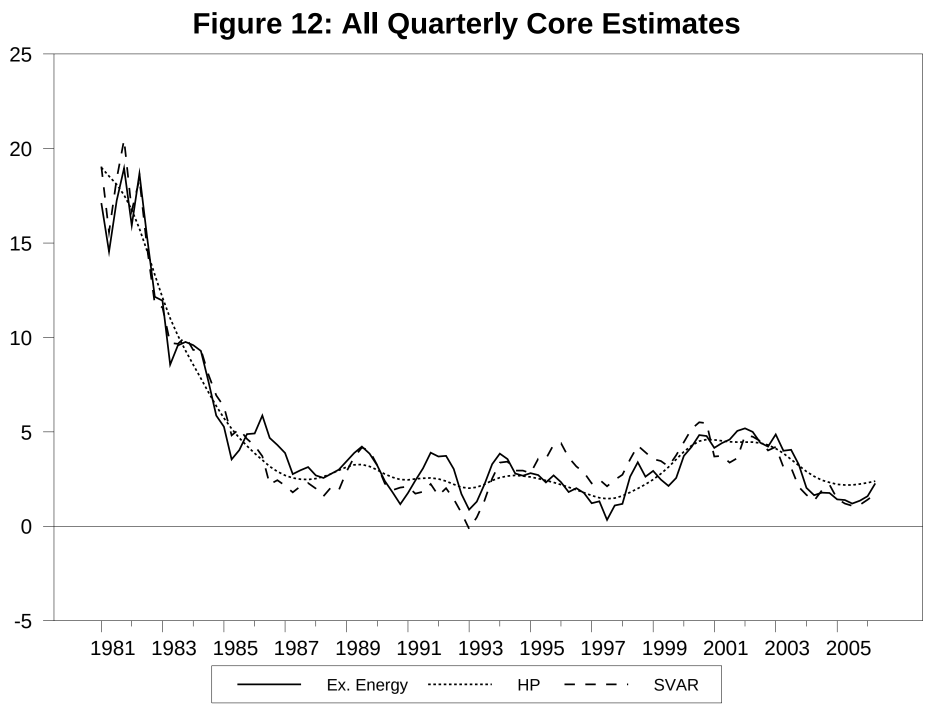


Figure 12: All Quarterly Core Estimates
25
20
15
10
5
0
-5
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
Ex. Energy
HP
SVAR

Figure 13: RMSE from Monthly Forecasts

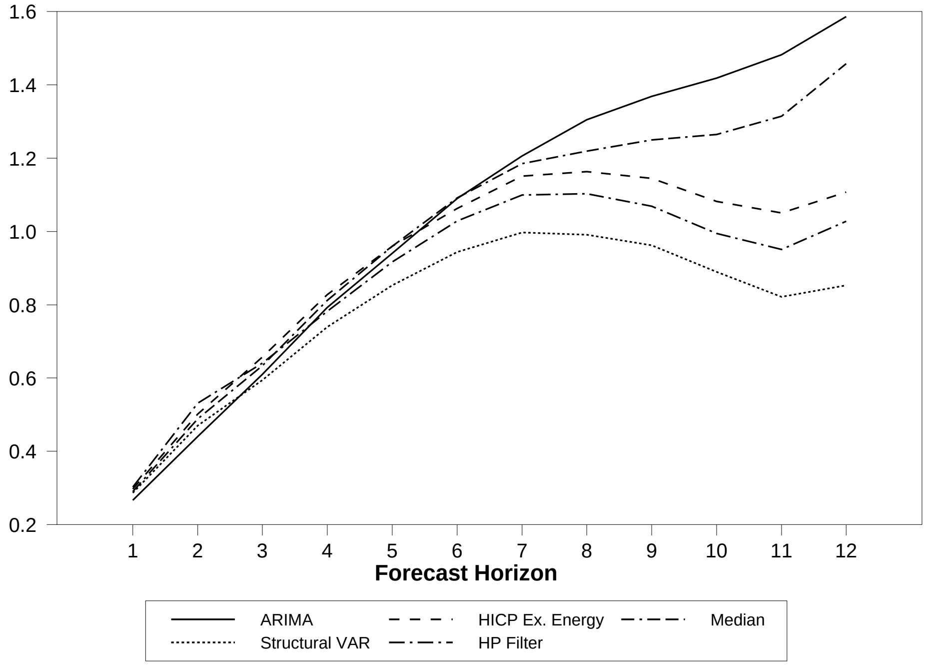
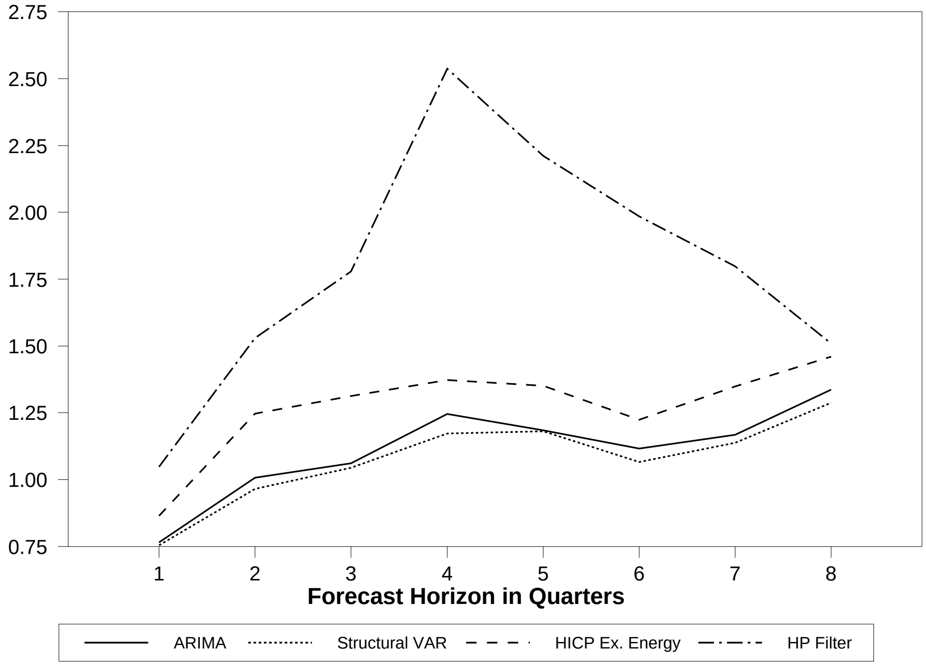


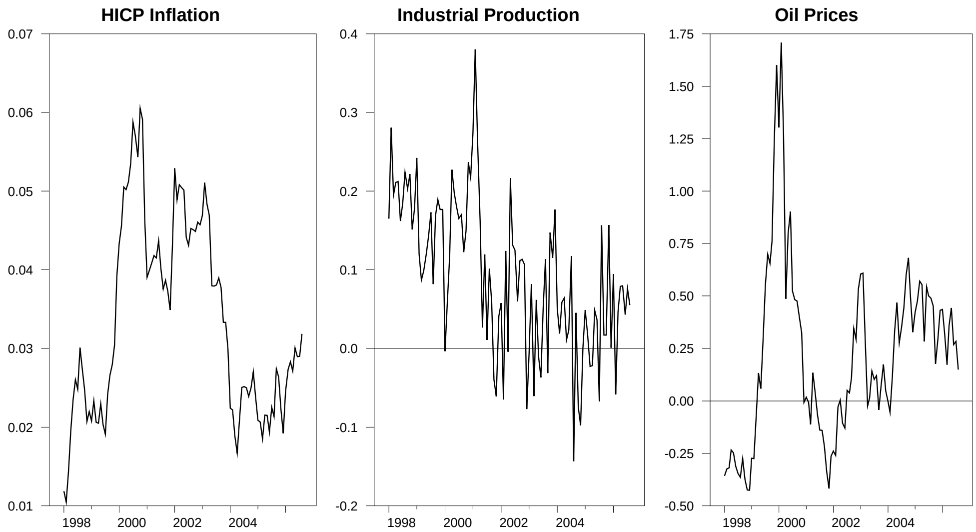
Figure 14: RMSE from Quarterly Forecasts



1.9 Appendix 1: Data Description

The data series are graphed and discussed in a bit more detail in this appendix. The monthly series are available over the period 1998M1-2006M8. This represents the most recent data available at the time of writing this study. The macro series used are inflation, industrial production growth and oil price inflation. Graphs of these monthly series are presented on separate graphs below given that the scales vary from series to series.

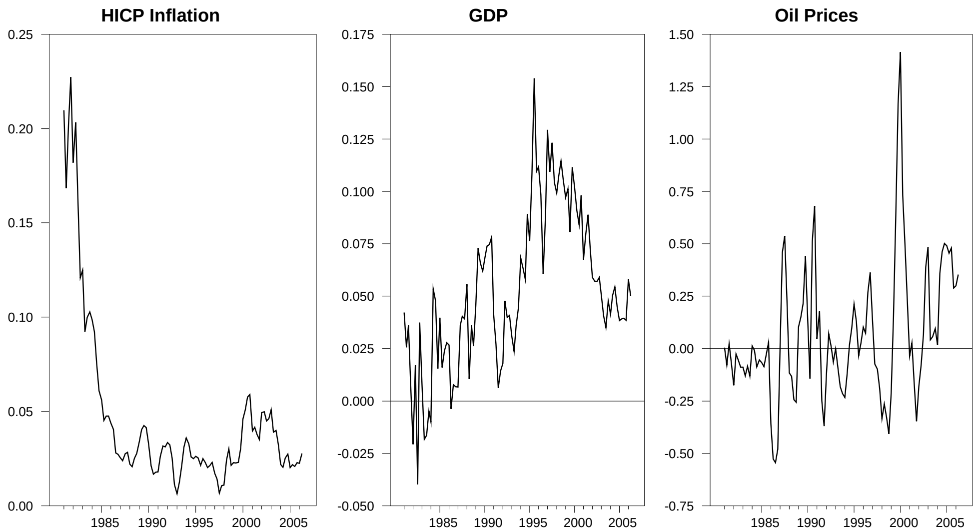
Figure A1: Monthly Variables



It's easy to see from the graphs why the unit root was not rejected for these series. Oil price inflation and consumer price inflation both spike in the earlier part of the sample and price inflation remains elevated for quite a while. With the exception of industrial production growth, the first differences of the other series are found to be stationary. Industrial production growth appears to have a higher mean in the early part of sample. Statistical tests confirm the structural break in the series and these types of breaks can incorrectly lead to the conclusion that a series is non-stationary. The first difference of this variable is stationary once the structural break is accounted for.

The quarterly series are graphed in the next figure and the data span here is considerably longer at 1980Q1-2006Q2. The high inflation rates carrying over from the oil price crisis of the late seventies is clearly evident at the start of the sample. The steady improvement in GDP growth over the sample is also evident and although the growth rate falls back at the end of the sample, the growth rates of about 5% at the end of the sample are still quite robust. Oil price inflation is volatile over the sample.

Figure A2: Quarterly Variables



The unit root tests for the quarterly inflation series finds that the series is stationary despite the early sample behaviour. The growth rate of GDP is found to be non-stationary. As with industrial production for the monthly data, this output series is also found to have a structural break and is stationary once this is accounted for. One question surrounding the data is whether is the monthly industrial production series is a good proxy for GDP growth. The series have similar stochastic properties. In addition, if the the monthly industrial production series is compacted to quarterly frequency through simple averaging, the quarterly industrial production growth series has a correlation coefficient of 0.69 with GDP growth. This suggests that it is a reasonable albeit imperfect proxy for GDP growth for monthly data. An alternative approach to arriving at a monthly output series would be to interpolate the GDP series to monthly frequency using something like the Chow-Lin procedure.

Chapter 2

A Critical Assessment of Existing Estimates of Core Inflation

Core inflation rates are widely calculated. The perceived benefit of core inflation rates is that they help to inform monetary policy. This is achieved by uncovering the underlying trend in inflation or by helping to forecast inflation. Studies which compare core inflation rates frequently assess candidate core rates on these two criteria. Using U.S. data, the two standard tests of core inflation - the ability to track trend inflation and the ability to forecast inflation - are applied to a more comprehensive set of core inflation rates than has been the case in the literature to date. Furthermore, the tests are applied in a more rigorous fashion. A key difference in this paper is the inclusion of benchmarks to the tests, which is non-standard in the literature. Two problems with core inflation rates emerge. Firstly, it is very difficult to distinguish between different core rates according to these tests, as they tend to perform to a very similar level. Secondly, once the benchmarks are introduced to the tests, the core inflation rates fail to outperform the benchmarks. This failure suggests that core inflation rates are of less practical usefulness than previously thought.

2.1 Introduction

The fundamental idea underlying the concept of core inflation is that inflation is ultimately determined by monetary growth, which should affect all prices in the economy equi-proportionately. Core inflation is then defined as the common element in all price changes. The concept is important because it provides a clear picture of the underlying trend in prices. This is in contrast to the actual inflation rate, which is inadequate for this purpose, as it is designed to measure changes in the cost of living. As such, the core rate is of particular importance in a policy context. Information regarding the true trend in prices is critical to policymakers given the long and variable lags between the implementation of monetary policy and its effect on inflation.

Like many other important economic concepts such as potential output or the NAIRU, core inflation is not an actual series and instead must be estimated. Its usefulness as a policy tool depends critically on the accuracy of the estimation method used to construct the core rate. Methods used to calculate core inflation include removing volatile items from the calculation of inflation, statistical filters, SVAR methods, trimmed means and factor models so there is a broad range of core inflation estimators. There is an existing literature that compares the relative merits of core inflation rates based on their policy usefulness. The two tests of policy usefulness most commonly used are the ability to track trend inflation and the ability to forecast actual inflation.

Using U.S. data, this paper proposes the most rigorous examination of core inflation estimates to date. The contribution of the paper is threefold. Firstly, in relation to the trend tracking test, the standard approach to date involves estimating the core rate and the trend over the full sample. This paper implements the test in a pseudo real time environment and so provides a more realistic assessment of the ability of core inflation rates to track trend inflation. Secondly, the forecast tests in the core inflation literature frequently omit a benchmark forecast from the comparison and instead only rank candidate core inflation rates. This is despite the convention in the forecasting literature of including a benchmark model. In this paper, benchmarks are included in both the forecasting tests and the trend tracking test. The introduction of a benchmark forecast to a core inflation paper is not novel. However, studies with benchmark forecasts have typically focused on a small number of core estimators and it has generally not been implemented for the US. In contrast, the introduction of the benchmark to the trend tracking test is novel. The final contribution of the paper is that the comparison exercise is the most exhaustive to date in terms of the range of core inflation estimators included. Most papers focus on a specific type of core inflation estimator whereas a number of core inflation estimators are included in this paper.

The paper finds that core inflation rates are no better at forecasting inflation or tracking trend inflation than the benchmarks included in the tests. In short, the benefit of core inflation rates to policymakers is overestimated. New uses of core inflation rates could exist. This paper suggests two alternative tests of core inflation, less stringent that those currently employed, but the performance of existing core inflation estimators is still relatively poor according to these alternative tests. It may also be possible to use existing estimators more efficiently. For example, although this paper shows that the standard forecasts of inflation based on an inflation gap fail to outperform a benchmark, other specifications or estimation techniques involving existing core rates might be found that could improve on the benchmark. The next section contains a literature review and highlights the contribution of the paper. Section 3 outlines the estimators used in the paper, including any issues in the estimation. Section 4 critically evaluates the performance of the core estimators and section 5 concludes the paper.

2.2 Literature Review

In tackling the issue of core inflation, the initial focus in the literature was simply to construct new estimates. A number of approaches were taken but these can generally be classified as either structural or statistical. The most basic statistical approaches simply involve excluding certain components, such as the volatile food and energy components. This type of core inflation rate is routinely calculated by national statistical agencies. More sophisticated techniques include statistical filters. The Hodrick-Prescott (HP) filter has been widely applied to economic time series, including inflation and provides one core estimate. The HP filter has been criticised in the past, particularly in relation to the well known end-point problem. Baxter and King (1999) propose an alternative filter, based on the spectral decomposition of a time series. It involves filtering parts of the series that lie between certain frequencies and this can be also used as a measure of core inflation.

Bryan et al (1997) propose the use of trimmed means as estimators of core inflation. Based on the notion that the headline rate can be significantly affected by large price changes in individual components, the trimmed means exclude these items and are considered robust to these outliers. Subsequent to their paper, trimmed mean estimates were calculated for a large number of countries. In this paper, we calculate trimmed means using two alternative weighting systems.

Persistence measures of core inflation can also be calculated. These measures are based on the persistence of the individual components that constitute the inflation rate. Persistence is estimated using an autoregressive model. Cutler (2001) applied this approach to UK data using only one lag for all series whereas Bilke and Stracca (2007) apply a similar approach to Euro Area data but measure persistence with the lag length determined using traditional lag selection tests. One of the core inflation measures examined in this paper is the Bilke and Stracca (2007) approach. This type of core inflation measure is rarely calculated so its inclusion in the comparison should shed some light on its relative merits.

The structural approach considered is the structural VAR as this is clearly the most prevalent structural approach to estimating core inflation. The methodology used is that proposed by Quah and Vahey (1995) with a standard long-run restriction. According to their approach, "inflation is assumed to be affected by two different types of shock, distinguished by their effect on output. The core inflation shock is output neutral after some fixed horizon whereas the non-core shock is allowed to influence output in the long-run." Following identically the method of Quah and Vahey (1995), a bivariate VAR is estimated using the assumption that the core shock is output neutral.

The final type of core inflation rate included in the paper is a dynamic factor model estimate. Factor models are used when analysing a large volume of data such as the individual price series that make up the overall inflation rate. Following the approach of Stock and Watson (2002), the factor model finds the common element in all these price changes. The benefit of this type of approach is that it takes time series information, cross-sectional information and frequency domain information into account.

The papers mentioned so far relate to the estimation of core inflation. Other papers in the literature aim to compare and assess various core inflation measures. This paper compares core inflation measures but considers a broader range of core inflation series than other papers in the literature. For example, Clark (2001) compares core inflation measures but concentrates chiefly on exclusion based statistical measures. In a study on German data, Landau (2000) includes the structural VAR but omits a number of important statistical estimators. Smith (2004) examines filters, trims and some exclusion measures, as do Rich and Steindel (2007). The scope of this paper includes all major estimation methods.

Many papers rank core inflation rates based on their ability to forecast actual inflation. Given the well documented difficulties associated with forecasting inflation it is somewhat surprising that this is such a popular yardstick. It is in some part due to the manner in which the forecast comparison exercises have been conducted. Although not an exhaustive list, Cogley (2002), Smith(2004), Clark (2001) and Rich and Steindel (2007) include only core inflation rates in the forecast comparison exercise using US data - there is no benchmark forecast included. The inclusion of a benchmark forecast is considered standard practice in the forecasting literature. Model forecasts are compared to forecasts from naive models, such as a no change forecast, in order to assess their forecasting ability. If the model forecast cannot beat the naive forecast, the model is of little worth for forecasting. This paper includes a benchmark in both the forecast test and the trend tracking test.

An additional improvement is also made to the trend tracking test. The trend is routinely defined as a centred moving average of inflation. The standard approach is to estimate the trend and the core rate using the full sample of data and then compare the two. Instead, we estimate the core and trend recursively as this more closely reflects the situation faced in practice. Although we stop short of conducting a full real time exercise, most core rates are based only on inflation data which are rarely revised.

We also consider two alternative metrics to gauge candidate core inflation series. The first criterion considered is the ability to predict changes in the direction of inflation. Although a poor predictor of the magnitude of inflation, core inflation may still be useful as a predictor of the direction of future changes. The second criterion that we examine is a measure of concordance, which has been used by McDermott and Scott (1999) in the business cycle literature. A key property of core inflation is to indicate whether there is excess inflationary pressure in the economy. If core inflation is above overall inflation, there is a negative "inflation gap". The ability of candidate core rates to measure this gap is captured by concordance, which is the degree to which core inflation series agree on the sign of the inflation gap. The performance of the core rates according to these tests do not suggest an alternative use for core inflation rates.

2.3 Calculation of Core Inflation

This section describes the construction of the core inflation measures. One issue of concern is the stationarity of the inflation rate. For some core measures, the unit root properties of inflation are irrelevant. This mainly applies to the statistical measures. The HP filter simply smoothes the inflation rate to get a core measure so the unit root issue is irrelevant. Similarly, trimmed means and the PCE excluding food and energy inflation rate both exclude some components of inflation. Once these items are excluded, the inflation rate is re-constructed. The unit root properties of inflation do not matter for this type of core inflation measure.

The paper also considers some time series methods to calculate core inflation, such as the persistence and SVAR measures, and the unit root properties of inflation take on more significance here. There is some doubt regarding the empirical unit root properties of inflation as the results can vary depending on the unit root test employed. Consequently, the SVAR model is estimated twice, first assuming inflation to be stationary and second time assuming a unit root. For reasons explained in the relevant section, the persistence measure is only estimated under the assumption that the component inflation rates are stationary.[1] For the bandpass filter, stationarity is also an issue. In this paper, we only apply the filter to the PCE inflation rate so this implicitly assumes the inflation rate to be stationary. The resulting core series has reasonable properties.

[1] First differencing non-stationary series and then applying the methodology did not result in a persistence measure that differed systematically from the first measure.

The main dataset used in the calculation of the core inflation rates is the Personal Consumption Expenditure (PCE) dataset from the National Income and Products Accounts (NIPA) tables. For some core measures, only the aggregate PCE inflation rate is needed. For other measures, a detailed breakdown of the PCE based on price indices for 206 separate items is used. This specific breakdown of the PCE together with the associated weights needs to be constructed manually from the data available on the website. Specifically, the series are taken from the underlying data which are available at bluehttp://www.bea.gov/national/nipaweb/nipa_underlying/SelectTable.asp. A cautionary note from the BEA warns that the underlying data may be of a lower quality than the data normally published. However, when the inflation rates of the 206 items were multiplied by the associated weights, it was possible to recover the aggregate PCE inflation rate with a very high level of precision, indicating that there are no quality issues with this part of the dataset. Quarterly data spanning 1960:1-2008:4 is used. For the structural VAR, data on real GDP over the same time period is also used.

2.3.1 Hodrick-Prescott Filter

The first estimate of core inflation used in the paper is the Hodrick-Prescott (HP) filter. The filter attributes a certain proportion of each shock hitting the series to a change in the trend of the series while the remainder is regarded as temporary noise. The wide use of HP filters in the profession and their ease of calculation warrants their inclusion in the study. Given the quarterly data used in the study, the standard value of 1600 is chosen for the smoothing parameter. The smoothed series is defined as core inflation. Figure 1 graphs the HP filter and it has the familiar properties.

2.3.2 PCE excluding Food and Energy

The inflation rate excluding food and energy was included as it is routinely computed by statistical agencies and is one of the most commonly referred to measures of core inflation. The idea is to exclude the items that are normally most volatile. A drawback to this measure is that food and energy are not always the most volatile components in the index. In addition, despite some volatility, they may contain some information regarding the core inflation rate which is lost by total exclusion. Figure 2 shows that this core estimator has been lower than actual inflation at the end of the sample given the high energy and food price inflation experienced in recent years.

2.3.3 Trimmed Mean

Trimmed Means are commonly constructed as measures of core inflation. The use of trimmed means is motivated by the leptokurtic distribution of individual price changes. This means that price change distributions generally have more extreme values than one would expect from a normal distribution and may be unduly influenced by these extreme values. By trimming the distribution, one removes the influence of these outliers and a more representative measure of the underlying inflation rate is obtained.

The standard approach to calculating a 10% trimmed mean is to order the inflation rates of the individual items from the largest to the smallest. Exclude the largest and smallest price changes (5% in each case for a 10% trim), re-scale the weights of the remaining items so that they again sum to one and calculate the inflation rate again as the weighted average of the remaining items. The problem with this sort of approach is that the weights are based on expenditure shares on a representative basket of goods, devised by statistical agencies to approximate changes in the cost of living. Among others, Wynne (1999) argues that there is no reason to believe that this weighting system should still be used when constructing a core inflation measure, which aims to capture the underlying trend in inflation rather than the cost of living. We argue the weights should be ignored for the following reason.

Consider the case where a 10% trimmed means is calculated using the 13-item breakdown of the PCE inflation rate. Assuming an asymmetric trim, the trimming operation results in just 1 of the 13 items being excluded so, in this case, only the most volatile item is removed. However, in the early part of the sample, the food item had a weight of about 25% and food is often one of the most volatile price indexes. Thus, to trim this inflation rate removes 25% of the index in terms of weights. To rescale the weights of the remaining items and call the resulting series the 10% trimmed mean is misleading. The severity of the problem is lessened when trims are applied to datasets with hundreds of items but the basic criticism still applies. For this reason, we prefer to trim the most volatile inflation rates and take a simple average of the remainder. We refer to the first approach as the weighted trim, as the weights are re-scaled following the trimming operation and refer to this second approach as the simple trim. In this application, the simple trim tends to perform better according to most criteria but only marginally. Figure 3 highlights that the weighted and simple trim have behaved quite similarly over the sample period.

2.3.4 Band Pass Filter

Following the methodology of Baxter and King (1995), a band-pass filter is applied to the PCE inflation rate to construct another core measure. Band-pass filters are based on a spectral decomposition of the time series and thus operates in the frequency domain of the series rather than the time domain. The spectral representation theorem states that a covariance stationary stochastic process can be expressed as a (infinite) weighted sum of periodic functions. It is the frequency domain analogue of Wold's representation theorem in the time domain. The periodic components are mutually orthogonal and have their own variance. The upshot of this is that we can isolate periodic components at specific frequencies.

The ability to isolate certain frequencies means that new series can be created by filtering out certain periodic components at specific frequencies. The implications in terms of constructing a core inflation measure are obvious. The noise component in the headline rate is defined as the high frequency component. By removing this high frequency component, we are left with an underlying series whose behaviour is driven by long-term trends. It is a more sophisticated approach to removing high frequency noise in comparison with the persistence approach but it also differs to the extent that it is applied directly to the PCE inflation rather than its component parts. Although there is clear potential to apply the filter to the components, it is applied to the aggregate as this is the convention in the literature and this study aims to assess the properties of the core rates, constructed as they are found in the literature. The filter is implemented so that components of the series with periodic fluctuations with a frequency of less than one quarter are filtered out. This removes the high frequency component of inflation. Once this high frequency component is removed, the underlying series is defined as core inflation. Figure 4 graphs the band-pass filter.

2.3.5 Structural VAR

A bivariate SVAR is also used in the paper to calculate another candidate for core inflation. This is the only structural estimate in the paper; the others are purely statistical in their construction. The variables included in the specification are the inflation rate and real GDP. This choice is based on a desire to produce the original SVAR core inflation rate, calculated in the seminal paper of Quah and Vahey (1995). In order to achieve structural identification, the standard restriction that the core inflation shock is output neutral in the long-run is imposed. This is consistent with the idea of a vertical long-run Philips curve and is a traditional identifying assumption in the application of long-run restrictions. Two core measures are calculated based on the assumption of a stationary and non-stationary inflation series. The two series are found to behave quite differently, as demonstrated in Figure 5.

2.3.6 Persistence Measure

Persistence measures of core inflation are amongst the least well-known and least widely implemented measures of core inflation. They are somewhat similar in spirit to the exclusion measures, such as inflation excluding food and energy. The exclusion measures exclude high variance components as they are considered to constitute noise. Advocates of the persistence approach prefer to classify noise as high frequency rather than high variance components. The idea is to increase the weight of the persistent components of inflation. This approach has some attractive features empirically. Given the possibility that some price series will exhibit both high variance and persistence, do we really want to exclude items based on variance only? The persistence measures are an attempt to address this short-coming in exclusion measures.

The persistence of a component is measured by estimating an AR model and ranking the magnitude of the autoregressive coefficients. The specific implementation has been approached in two ways in the literature. Cutler (2001) recommends estimating the following AR model using monthly data and annual inflation rates:

$$\pi_{i,t} = \alpha_{i,t} + \rho_{i,t}\pi_{i,t-12} + \epsilon_{i,t} \tag{2.3.1}$$

The subscript i is used to index across the various components. The estimated magnitude of the autoregessive coefficient is the persistence estimate. If this coefficient is negative, it is evidence of very fast mean reversion and the item in question is given a zero weight in the persistence measure. For the other components, their weight is proportional to the magnitude of the autoregressive coefficient. This approach is somewhat restrictive in terms of the specification of the autoregressive model. The approach implemented in this paper follows that of Bilke and Stracca (2007), who estimate a model of the form:

$$\pi_{i,t} = \alpha_{i,t} + \sum_{j=1}^{q_i} \rho_{i,j}\pi_{i,t-j} + \epsilon_{i,t} \tag{2.3.2}$$

In this case, the lag length of the autoregressive model is chosen according to the Schwartz information criteria. Lag lengths up to twelve lags are considered although in most cases, the lag length chosen was quite short - the average lag length was just over 2. The persistence measure is the sum of the estimated autoregressive components. As in Cutler (2001), items with negative sums are given a zero weight. Following Bilke and Stracca (2007), the inflation rates are re-weighted in proportion to the magnitude of the summed AR weights. The coefficients are not estimated in a time-varying manner in the traditional sense although the fact that these measures are evaluated recursively means that they will change slightly over time as the sample changes.

The papers in this area are unclear as to whether each series should be tested for a unit root. When there are a large number of series available, clearly some will have unit roots while others will not. However, as the method involves a re-weighting of inflation rates, it seems logically inconsistent if some items are weighted based on the persistence of inflation rates while others are weighted on the persistence of the first difference of inflation rates. For this reason, we apply the AR model to the inflation rates only, which appears to be the standard approach in the literature. Figure 6 shows that the persistence measure has tracked actual inflation quite closely over the sample.

2.3.7 Exponential Smoother

Cogley's (2002) exponential smoother also aims to capture persistent movements in inflation. However, this persistence is motivated in terms of the behaviour of central banks. The idea is that shifts in mean inflation arising from changes in policy rules are the main source of inflation persistence and core inflation should be designed to adapt to these changes. The exponential smoother is designed to measure changes in mean inflation, whereby the mean of inflation is updated based on new data according to a constant gain algorithm. This updating rule corresponds to simple exponential smoothing, which is a one-sided geometric distributed lag of past inflation:

$$\pi_t^* = g_0 \sum_j (1 - g_0)^j \, \pi_{t-j} \tag{2.3.3}$$

where π^* is the exponential smoother, g_0 is the gain which is calibrated based on the values suggested by Cogley (2002) and π_t is actual inflation. As with other estimators which aim to isolate the persistent elements of inflation, this filter removes the high frequency component. It differs from the persistence estimator in the sense that it is applied directly to the PCE inflation rate and it differs from the HP and bandpass filters to the extent that it is a one-sided filter and so does not suffer from an end-point problem. Figure 7 graphs the exponential smoother and it has the characteristic properties of this type of filter.

2.3.8 Factor Model

The factor model used follows the approach of Stock and Watson (2002) in that we estimate a static representation of a dynamic factor model. This type of model can be estimated using principal components.[2] Each individual inflation rate is assumed to be driven by a small number of common factors and an idiosyncratic error. The common factor is based on a decomposition of the covariance matrix of the standardized inflation series. In particular, the covariance matrix is decomposed in terms of its eigenvalues and eigenvectors. The eigenvector has as many elements as there are inflation series and it provides the weights need to construct the factor. The factor is a linear combination of the inflation rates which explains the largest possible amount of variance and covariance in the underlying inflation series. In this way, the factor is capturing the common dynamics in the data and is often loosely term the 'common component' in the core inflation literature.

$$\pi_{i,t} = \Lambda_i F_t + \epsilon_{i,t} \tag{2.3.4}$$

Each inflation rate is related to the factors with unique factor loadings, Λ_i. This means that each individual inflation rate can be expressed in terms of the common component and the idiosyncratic component. Core inflation is that part of overall inflation that is related to the common factor - it removes the idiosyncratic component. Prior to estimation, all inflation rates must be transformed to ensure stationarity. By definition, the resulting core estimator is also stationary. The use of factor models is most common in the pure forecasting literature but it has been fairly widely applied in the core inflation literature also. It represents a hybrid of the statistical approaches

[2]See Stock and Watson for technical details.

in the sense that both time series and cross sectional information is used in its construction. Figure 7 shows the factor estimate of core inflation and it's notable that this estimate was considerably higher than actual inflation during the first oil price crisis.

2.4 Comparison of Core Measures

Having outlined the core measures included in the paper, we now begin the evaluation process and this section contains the key contributions of the paper. The two standard tests are the ability to track trend inflation and the ability to forecast inflation. Improvements are made to these two standard tests and the tests are applied with the most comprehensive set of core estimators to date. Additional tests not normally found in this literature are also applied. The following results, therefore, provide the most realistic appraisal of the practical usefulness of U.S. core inflation estimators.

2.4.1 Summary Statistics

To begin the analysis of the various core measures, a couple of basic summary statistics are presented for each core series. Although these are the most basic statistics for any series, it is often argued that they are especially important in the core inflation context. In terms of the mean of the series, one would expect a core inflation rate to have a similar mean to the headline inflation rate when considered over a long time span. If core inflation and actual inflation have significantly different means over a sustained period, the core measure is systematically divergent from the headline rate. Clark (2001) cites similarity of means as one criterion to assess the ability of a core measure to track the trend in inflation as "policymakers and other analysts prefer a measure of core inflation that neither understates nor overstates the long-term trend rate of price change". The importance of the standard deviation lies in the fact some of the core measures are constructed on the basis that volatile components are excluded. Thus, once volatile components are excluded, the resulting series should be less volatile. In this section, we examine the summary statistics of the core measures to see if this is a valid means to discriminate between candidate core series.

Table 1 presents summary statistics for the PCE inflation rate and the core inflation series calculated over the period 1963Q2 - 2008Q4. The PCE inflation rate has a mean of 3.79% over the sample and most of the core measures have a mean which is similar to this. Statistical series generally perform strongly on this criterion. The HP filter posts a mean inflation rate of 3.80%. The I(0) SVAR, the persistence measure and the factor model estimate all have means which differ from the PCE mean by less than 0.03.% The band pass filter has a mean of 3.75%, again quite similar to the PCE inflation rate. As all core inflation rates have very similar means over such a long sample, this is not a suitable statistic to choose the best measure.

A second summary statistic often cited in relation to core inflation rates is their variance or standard deviation. Certain core estimators are designed to remove high frequency noise and according to this criterion, good estimators should have a lower variance or standard deviation than the actual inflation rate. The second column of the table shows the standard deviation of inflation and the core measures. The HP filter, BP filter, SVAR and exponential smoother all do well on this metric. However, the difference in volatility is not of a sufficient magnitude to meaningfully discriminate between these core measures. In addition, seven core series have a standard deviation between 1.98, which is the lowest value, and 2.10. The table also presents the correlation of the core measures with the PCE inflation rate and their correlation with a centred moving average of inflation. Again, all series are highly correlated with the PCE inflation rate. They are also highly correlated with the centred moving average so the summary statistics do not provide a basis for choosing amongst core inflation rates.

2.4.2 Tracking Trend Inflation

The ability to track trend inflation is often considered a key property of a good core inflation rate. Bryan et al (1997), Cecchetti (1997) and Clark (2001) all define the trend in inflation as a Centred Moving Average (CMA) of the headline inflation rate and this is the standard definition of the trend in inflation in the literature. A centered moving average uses past and future values of the series when estimating the trend. One shortcoming in the literature is that a core inflation rate constructed using the full sample of data is used as the basis for comparison with the CMA trend. The historical estimates of the trend from the full sample core inflation rates will differ from those that would have been available in reality, when the full sample of data were not available. In this paper, we construct core inflation rates recursively to more accurately reflect the situation faced in real time. This has two important benefits. Firstly, a common criticism of econometric estimates of the core rate is that it changes every quarter as the model is re-estimated with additional data. By estimating the core inflation rates recursively, we construct the core measure that would have been available to policymakers at each point in time. The core series are estimated recursively over 1960Q1-1989Q1 in the first step and over 1960:1-2008Q4 in the last step, which represents a period of twenty years. The core inflation estimate for the current quarter of each recursive step is compared with the CMA trend, as the policymaker is generally most interested in the estimate of the trend for the current quarter when setting policy. The second benefit of the recursive estimation strategy is that it takes account of the end-point problem with statistical filters and so provides a more realistic indication of their ability to track trend inflation.

Table 2 presents the results of the recursive trend tracking ability of the core inflation series. The first column shows the correlation of the core series with a nine quarter CMA, which is used as the estimate of the trend. When we compare the correlations in this column to those in the last column of Table 1, which represented correlations with core rates estimated over the full sample, we can see that in most cases, there is a slight decline in the core series' ability to track trend inflation once the exercise is performed recursively.[3] The correlation with the CMA trend declines significantly for the stationary SVAR measure of core inflation although the persistence estimate actually increases its ability to track the trend. The HP filter has a high correlation with the trend even though it is reduced to a one-sided filter at the end of the sample. There are seven core series which have a correlation of 0.90 or higher with the CMA trend so this criterion again leaves little to choose amongst the core series.

Although not reported in Table 1, the correlation among the series is also quite high. It is possible to create a correlation matrix for the core measures and then average by column. This gives the average correlation between one core measure and all its comparators. The range of values for these averages is between 0.88 for the SVAR I(1) measure to 0.95 for the weighted trim. The tight range of values might suggest that these core rates are all catching the core inflation quite well. However, we will re-examine this issue in the context of another bivariate statistic following the forecasting exercises.

In general, trend tracking tests have no benchmark. In this paper, a benchmark defined as a five quarter moving average is introduced. This benchmark is adopted as it is trivial to compute and is a one-sided filter. Thus, if we choose to define the trend as a 9-quarter centred moving average, we can choose a benchmark defined as the currently available part of that centred moving average. The last row of the table shows the correlation of this 5-quarter moving average with the centred moving average trend. This benchmark correlates with the trend just as highly as any core series. Therefore, the core inflation rates perform no better than this benchmark.

A second way to measure the ability to track a CMA trend is to calculate average deviations from the trend. The second column of the table presents the Relative Mean Absolute Error (RMAE), calculated over the recursive sample using the formula:

$$RMAE_i = \frac{\sum \|(\pi_{core,i} - CMA)/CMA\|}{n} \tag{2.4.5}$$

This calculates the absolute difference between the core series and the trend as a fraction of the trend and finds the average. The persistence series performs well according with a value 0.09 for this statistic. However, the 5-quarter moving average has an average error of 0.10 and the trims both have errors of 0.11, as does the band pass filter. Thus, the range of values for this statistic is quite narrow and no core series dominates. On the basis of the two trend tracking tests, a number of series core inflation series perform quite well but there is no clear front-runner. In addition, the

[3]The correlations from the two tables are not strictly comparable as the sample period is considerably shorter for the recursively estimated series in Table 2.

benchmark does just as good a job of tracking the trend when the standard definition of a centred moving average is adopted. Consequently, the ability to track trend inflation, whether through correlations with the trend or deviations from it, is not a fruitful avenue in terms of ranking core inflation series. More importantly, core inflation rates are no more useful than a lagged moving average in terms of tracking a trend when the trend is defined as a centred moving average.

2.4.3 Forecasting Headline Inflation

The ability to forecast inflation is cited as a key indicator of the policy usefulness of core inflation rates. In this section, competing measures of core inflation are ranked according to their ability to forecast the headline inflation rate and a benchmark forecast is included in the analysis. Forecasts are constucted using the following regression, which is the standard forecasting equation in the literature:

$$\pi_{t+h} - \pi_t = \alpha + \beta \left(\Pi_t - \pi_t \right) + v_t \tag{2.4.6}$$

where π_t is the inflation rate at time t and Π_t is core inflation. The left hand side of the equation is the difference between headline inflation today and headline inflation h periods in the future. On the right hand side, the term in brackets is the difference between core inflation and headline inflation. The basic premise of this forecasting regression is that the difference between headline inflation and core inflation today has predictive power for headline inflation tomorrow. In particular, if there is a large divergence between headline inflation and core inflation, you would expect headline inflation to move back towards core inflation because core inflation is a measure of the general trend in inflation.

The regression computes a forecast over a fixed horizon. For example, using quarterly data and setting $h = 8$ would yield a forecast of headline inflation eight quarters in the future but would not forecast inflation in the intervening periods. In order to get a continuous forecast to the end of the forecasting horizon, eight quarters in this paper, eight regressions of the type above are estimated setting $h = 1...8$. Each candidate core inflation rate is put in the regression equation above and forecasts of the headline inflation rate are generated. A "no change" benchmark forecast is used to compare the performance of the core series. Under this scenario, if inflation is 4 per cent in 2000Q1, the forecast for year-on-year inflation for each quarter in the forecast horizon 2000Q2-2002q2 is also 4 per cent.

The quarterly forecasts are performed on a recursive basis, with one observation added to the sample each time. In the first recursive step, estimates of core inflation are calculated over the sample 1960Q1-2000Q1 and forecasts are performed up to 2002Q2. The process is repeated adding one observation each time so by the end of the final estimation period, there are 28 sets of forecasts for each core estimation method. The forecasting exercise is repeated using only data from the start of the Great Moderation period. In these short sample estimates, the first recursive series are estimated from 1982Q1-2000Q2. Although 1985 is generally accepted as the beginning of the Great Moderation, we choose a period a few years earlier to begin estimation in order to allow more degrees of freedom in the estimation of the econometric series, particularly the SVAR measures. The forecast periods and number of recursive estimates are identical in both forecast exercises; only the estimation period changes. However, we also conduct a third forecast exercise over the full sample when the number of recursive steps is doubled to 56 as a robustness check. In this instance, the first estimation period runs from 1960Q1-1993Q1.

The results of the full sample forecasting exercise with 28 recursive steps are presented in Table 3. The numbers in the table are the ratios of the RMSE from the regression forecasts to the no change benchmark. A value less than one indicates that it is more accurate to forecast inflation using the regression forecast. From Table 3, we can see that the core inflation rates perform very poorly in terms of forecasting. The SVAR, where inflation is assumed to be I(1), has the best forecasting power. It is much more accurate than any other core measure, particularly at longer forecast horizons. However, it is less accurate than the no change forecast except for quarters five and six. Even then, the improvement in forecast power relative to the benchmark is marginal. The short sample estimates in Table 4 paint a similar picture. Table 5, which presents the full sample estimates with additional recursive steps, again shows that no core rate outperforms the benchmark. As this has the largest number of recursive steps, the results of this exercise are potentially the most robust. For this reason, formal forecast comparison tests are performed on the forecasts in this table. The Diebold-Mariano (1995) test of equal predictive ability is almost universally rejected. This indicates that the core inflation based forecasts are statistically inferior to the no change forecast. The only exceptions are the forecasts of quarter 1 and 2 from the HP filter. The systematic failure of core inflation regressions to beat a naive benchmark indicates that core inflation rates are not a useful tool in terms of forecasting PCE inflation.

2.4.4 Directional Forecasting

Although core inflation rates do a poor job of forecasting the magnitude of inflation, perhaps they are more suitable to predicting changes in the direction of inflation. Taking the forecasts from the previous section, they are evaluated according to their ability to correctly forecast the direction of the change in inflation four quarters ahead and eight quarters ahead. The forecasts are available over the full and short sample with 28 recursive steps and over the full sample with 56 steps. There is no benchmark per se in this exercise although one would wish that the forecasts would beat a coin toss so that the correct direction is forecast at least 50% of the time. The results presented in Table 6 give the percentage of times that the models correctly forecast the direction of change in inflation.

The core inflation rates do not generally perform well according to this statistic. If we look at the first two columns of the table, which represent the full sample estimates with 28 steps, the I(1) SVAR and the simple trim are the only two series to correctly forecast the direction of the change in inflation more than 50% of the time over both four and eight quarters. Columns three and four show the short sample results. The I(0) VAR correctly predicts the direction of change 64% of the time four quarters ahead while the excluding food and energy series does well for the eight quarter forecast. The factor model beats a coin flip for both horizons. For the full sample results with 56 steps, the I(1) SVAR is the only core rate with forecast accuracy greater than 50% at either four quarters or eight quarters. Taking the results as a whole, the failure of any core rate to systematically (i.e. across forecast exercises) beat a coin flip in terms of directional forecasting highlights major shortcomings in core rates as forecast tools. However, the I(1) SVAR is a front-runner in this exercise as it beats a coin flip in the two full sample exercises.

2.4.5 Concordance

Concordance is a broad measure of the degree to which the various core inflation rates agree with each other in terms of whether core inflation is above or below actual inflation. For example, if one core measure shows core inflation to be above actual inflation but all the others show it to be below actual inflation, one would conclude that it is below on the balance of evidence. A concordance measure puts this type of logic on a firmer statistical footing. In this context, the concordance statistic is a bivariate statistic that measures the degree to which two core inflation rates agree that core inflation is above/below the headline rate. More specifically, it measures the proportion of the time that two series are in the same state. If we define an inflation gap for each core series as the difference between the candidate core measure and headline inflation, we can define a corresponding series $S_{i,t}$ to be equal to 1 when the gap measure is positive and equal to 0 when the gap measure is negative, where the subscript i is an index over the different core inflation series. The degree of concordance for a pair of gap measures is then calculated as:

$$C_{i,j} = T^{-1} \sum \{(S_{i,t}.S_{j,t}) + (1 - S_{i,t})(1 - S_{j,t})\} \tag{2.4.7}$$

By construction, the value of the concordance statistic is bounded between zero and one. A value of 0.5 between two core series means that, 50% of the time, the sign of the inflation gap is the same when calculated using both core inflation rates. The concordance statistics are presented in Table 7. The core inflation rate with the highest average concordance is the exponential smoother. On average, it is in agreement with the other core inflation rates 71% of the time regarding the sign of the inflation gap. The excluding food and energy measure also performs well with average concordance of 70%. The I(1) SVAR has the least satisfactory performance according to this statistic.

Although the results appear reasonable here, there are also difficulties with this statistic in terms of ranking core inflation rates. The range of values for the statistic is again quite tight with five core inflation rate scoring between 0.66 and 0.71. The concordance statistic does not separate the different core inflation rates any clearer than the trend tracking statistic. Also, following the poor results of the directional forecasting exercise, one has to question whether any core rate is consistently measuring excess inflationary pressure in the economy.

2.5 Summary and Conclusions

The implementation of effective monetary policy requires an accurate assessment of the rate of core inflation in an economy. Like other important concepts such as potential output and the NAIRU, the core inflation rate is not an actual series and instead must be estimated. This paper conducts the most rigorous and comprehensive analysis of existing estimates of core inflation to date. There are other papers of this variety in the literature but these often focus on a specific type of core inflation estimator. This paper compares all major estimation methods. The exercise is conducted for the US and improvements are made to the standard comparison tests. In addition, extra tests not generally used in this literature are also applied to the core inflation rates. Two problems emerge in the comparison exercise.

Firstly, the candidate core inflation rates are very difficult to separate according to the comparison tests as a large number of estimators generally perform to a very similar level. This makes it very difficult to rank the core inflation rates. The two standard tests of core inflation are its ability to track trend inflation and its ability to forecast future inflation. Comparisons are mostly conducted just amongst the core rates. When simple benchmarks are included, no core rate can outperform the benchmark in either test. This calls into question the usefulness of existing core inflation measures. Additional tests not featured in the literature are also examined but the performance of existing core inflation estimates is still relatively poor.

As the literature has not highlighted these shortcoming of core inflation rates to date, future work is needed to determine if these results are specific to this dataset or perhaps specific to the US. It is difficult to foresee how the trend tracking ability of core inflation rates will compare in other studies. However, the general difficulties in forecasting US inflation in the post Moderation period suggest that the forecasting results are unlikely to be overturned for US data although factor model core estimators have demonstrated good forecasting properties for other countries. Also, given the wide variety of increasingly sophisticated techniques, it seems unlikely that forecasting U.S. inflation using OLS on an inflation gap will prove the best approach.

The results of the first chapter found core inflation rates to be useful. However, once the more stringent criteria are adopted in terms of including standard benchmarks, this result is overturned for the US. There are a large number of papers in the literature which follow the approach taken in the first chapter. This leads us to question to what extent the results of other papers in the literature, not just those using US data but also those based on data outside the US, would be overturned should these more realistic tests be used.

2.6 Figures and Tables

Figure 1: HP Filtered Inflation Rate

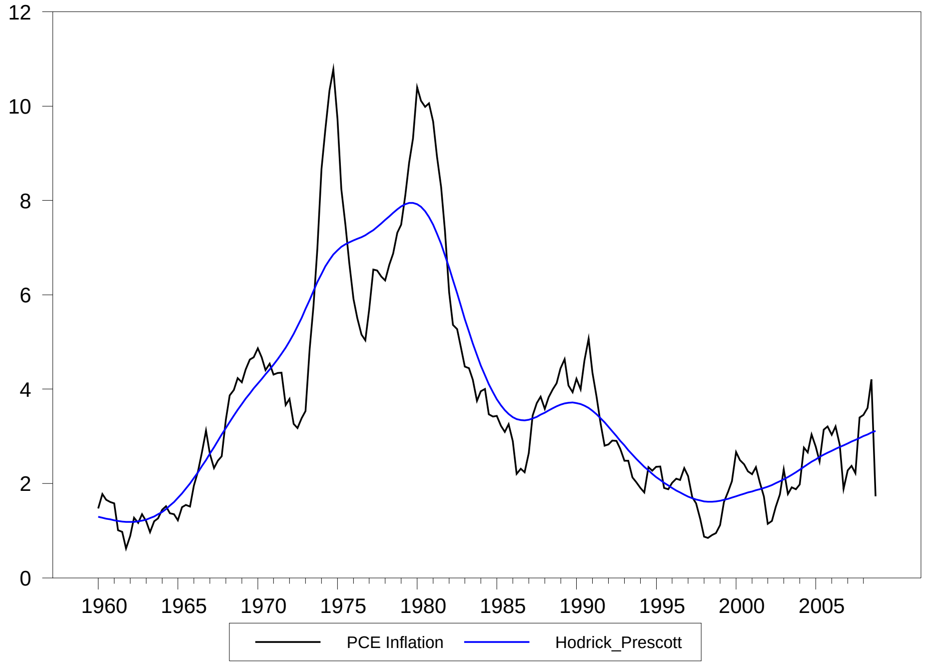


Figure 2: Inflation and Inflation excluding Energy and Food

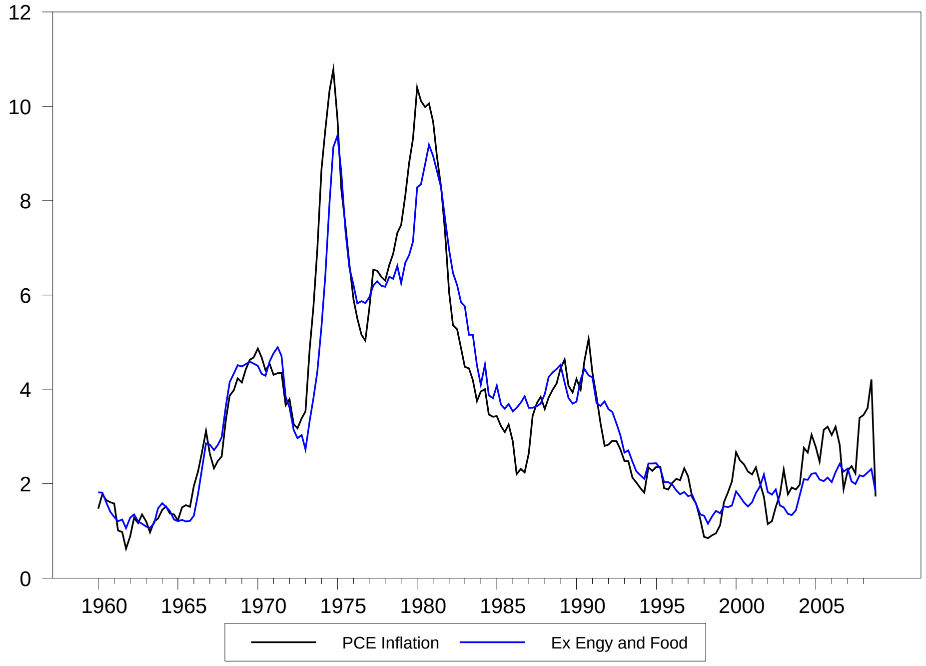


Figure 3: Inflation with Simple and Weighted Trimmed Means

Calculated using all 206 series

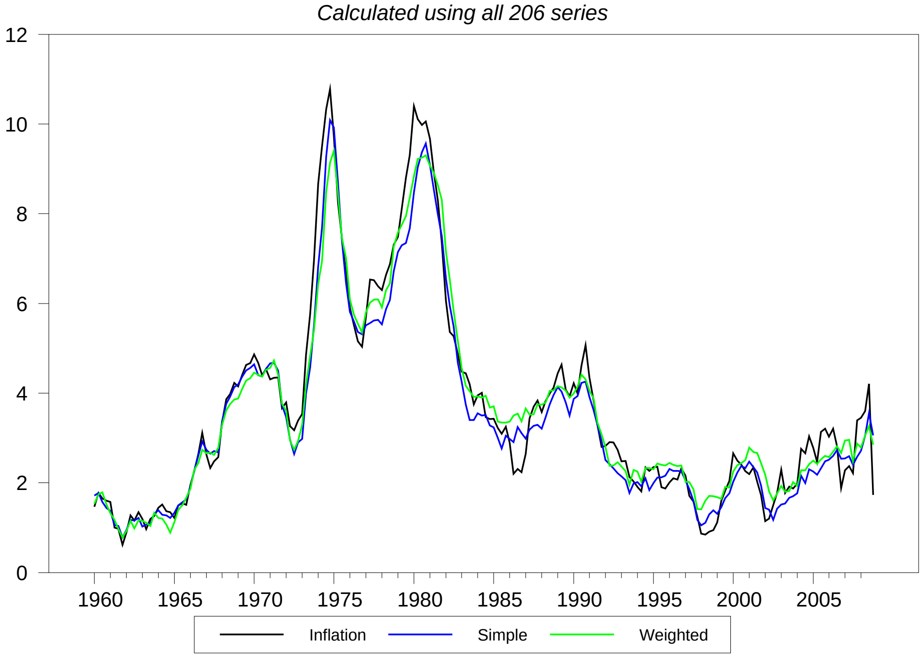


Figure 4: Inflation and Band Pass Filter

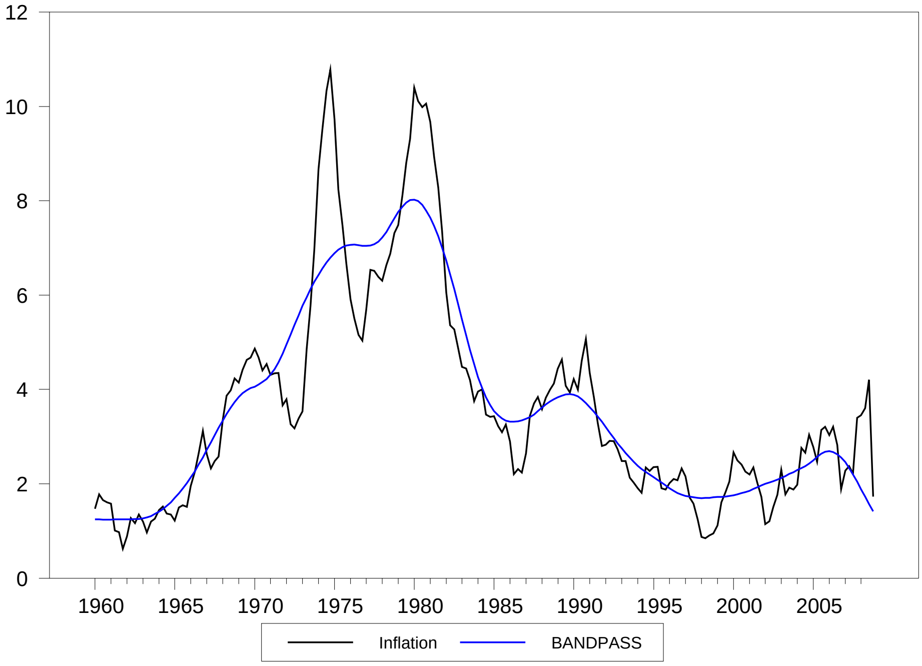


Figure 5: Inflation and SVAR Estimates

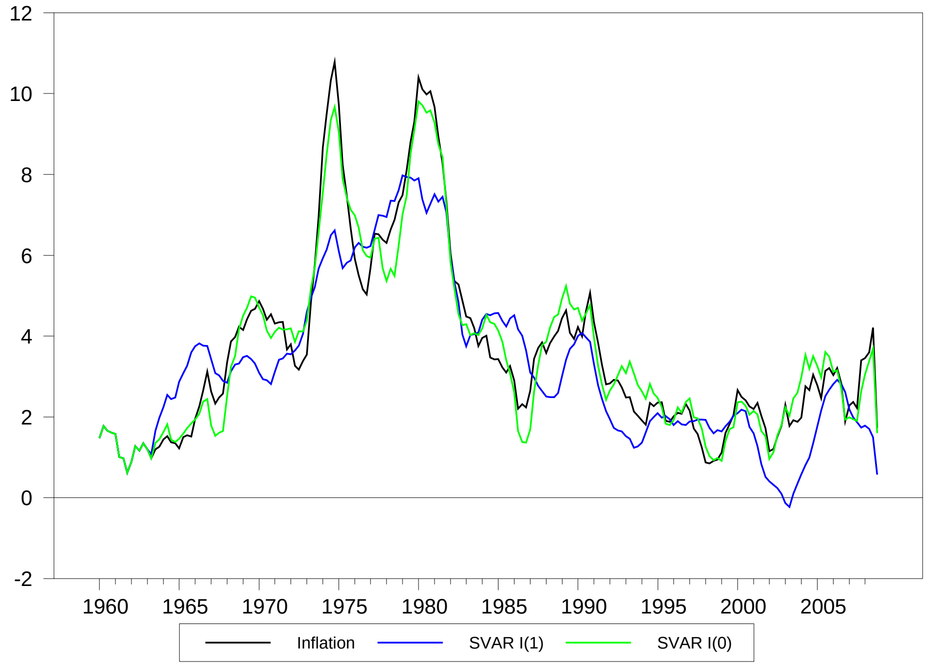


Figure 6: Inflation and Persistence Series

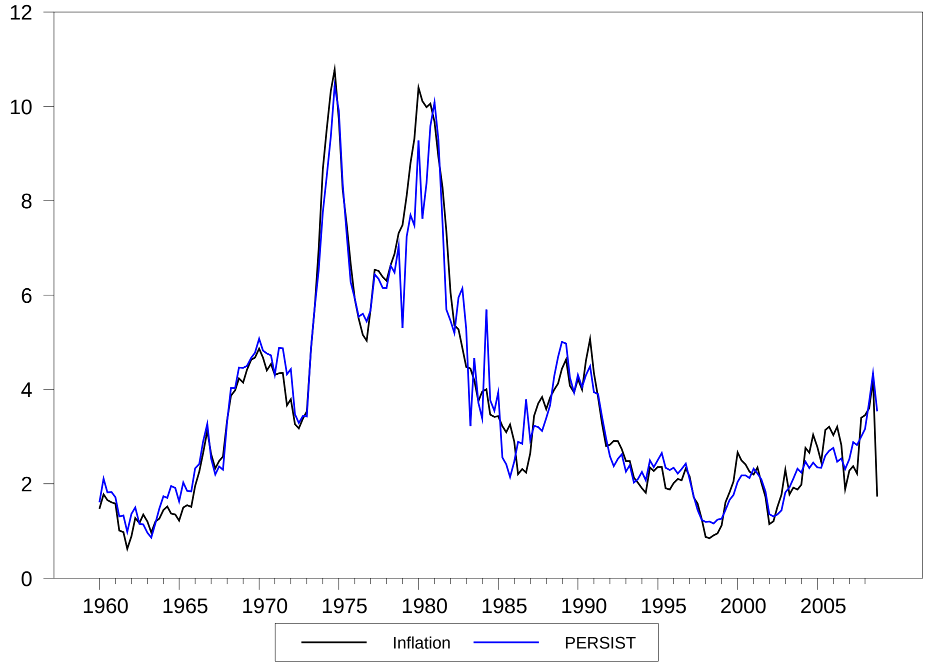


Figure 7: Inflation and Exponential Smoother

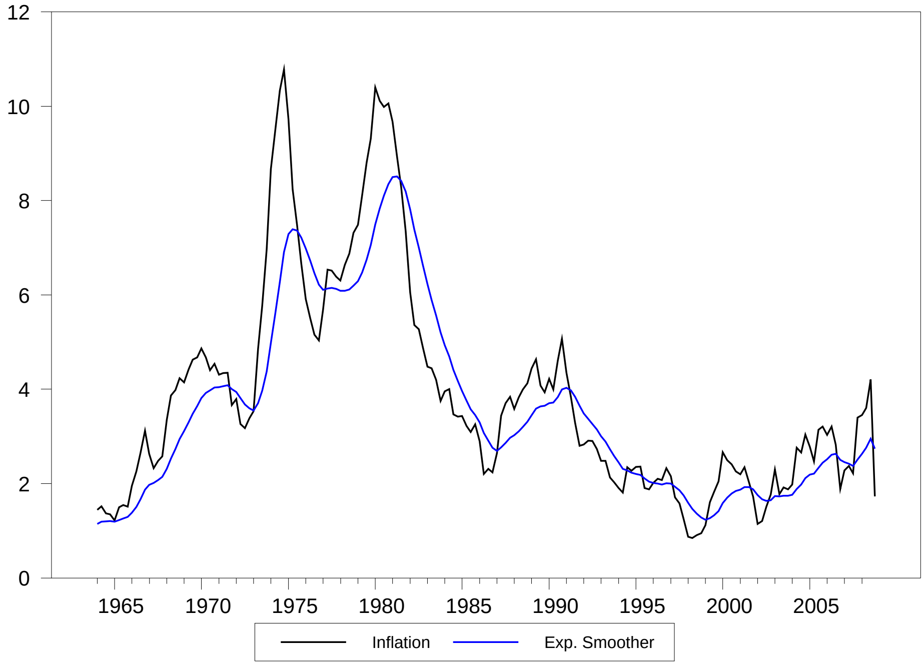


Figure 8: Inflation and Factor Model Core Estimate

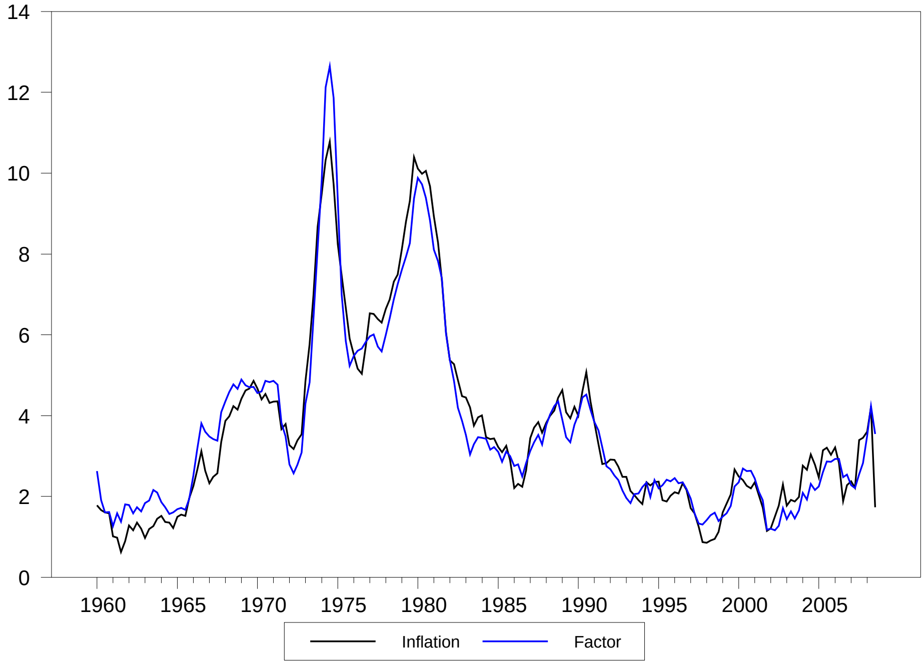


Table 1: Summary Statistics for Core Measures

Core Measure	Mean	Std. Dev.	Corr. PCE	Corr. CMA
PCE	3.79	2.31	1.00	NA
HP FILTER	3.80	2.00	0.91	0.96
EXC. FOOD & ENERGY	3.64	2.07	0.94	0.95
TRIM	3.57	2.10	0.98	0.98
WTRIM	3.74	2.07	0.97	0.98
SVAR I(1)	3.37	2.00	0.86	0.89
SVAR I(0)	3.77	2.19	0.98	0.96
PERSIST	3.76	2.11	0.97	0.93
BAND PASS	3.75	2.00	0.91	0.96
EXP. SMOOTH	3.51	1.98	0.90	0.94
FACTOR	3.77	2.27	0.97	0.95

Table 2: Recursive Trend Tracking Test

Core Measure	Corr CMA	Deviations
HP FILTER	0.92	0.13
EXC. FOOD & ENERGY	0.90	0.13
TRIM	0.94	0.11
WTRIM	0.93	0.11
SVAR I(1)	0.87	0.14
SVAR I(0)	0.84	0.16
PERSIST	0.95	0.09
BAND PASS	0.91	0.11
EXP. SMOOTH	0.95	0.13
FACTOR	0.90	0.14
MOV5	0.95	0.10

Note: The table shows the ability of each core measure to track the trend based on correlation to or deviations from the trend.

Table 3: RMSE from Full Sample Inflation Forecasts

Horizon	**Forecast Method**				
Quarters	SVAR I(1)	SVAR I(0)	HP Filter	BP Filter	EXP Smooth
1	1.01	1.07	1.04	1.07	1.12
2	1.01	1.15	1.11	1.20	1.26
3	1.00	1.22	1.22	1.36	1.41
4	1.00	1.24	1.35	1.49	1.53
5	0.98	1.30	1.47	1.63	1.64
6	0.99	1.35	1.58	1.73	1.72
7	1.06	1.33	1.62	1.74	1.71
8	1.05	1.33	1.66	1.80	1.72

	Forecast Method				
Quarters	Persistence	Ex. Food Engy	Trim	WTrim	Factor
1	1.06	1.23	1.19	1.20	1.01
2	1.14	1.48	1.37	1.43	1.04
3	1.22	1.70	1.49	1.62	1.08
4	1.28	1.81	1.57	1.76	1.13
5	1.29	1.96	1.60	1.89	1.21
6	1.24	1.99	1.56	1.90	1.28
7	1.21	1.88	1.48	1.82	1.32
8	1.19	1.87	1.38	1.79	1.39

Note: The table shows the ratio of the RMSE from a regression with the named core inflation rate to the "no change" forecast. A value less than one signifies a lower forecast error than the benchmark forecast. First estimation period: 1960:1-2000:1

Table 4: **RMSE from Post Moderation Sample**

Horizon	Forecast Method				
Quarters	SVAR I(1)	SVAR I(0)	HP Filter	BP Filter	EXP Smooth
1	1.02	1.04	1.12	1.20	1.17
2	1.04	1.05	1.27	1.44	1.34
3	1.04	1.03	1.41	1.64	1.48
4	1.02	1.03	1.56	1.78	1.59
5	1.01	1.02	1.70	1.92	1.69
6	1.02	1.02	1.80	2.00	1.75
7	1.03	1.04	1.85	2.00	1.74
8	1.03	1.04	1.96	2.13	1.81

Horizon	Forecast Method				
Quarters	Persistence	Ex. Food Engy	Trim	WTrim	Factor
1	1.02	1.16	1.10	1.12	1.09
2	1.04	1.38	1.23	1.28	1.21
3	1.04	1.61	1.30	1.37	1.27
4	1.04	1.77	1.38	1.46	1.34
5	1.03	1.92	1.40	1.51	1.28
6	1.04	1.95	1.38	1.48	1.19
7	1.02	1.86	1.38	1.46	1.16
8	1.05	1.91	1.40	1.50	1.10

Note: The table shows the ratio of the RMSE from a regression with the named core inflation rate to the "no change" forecast. A value less than one signifies a lower forecast error than the benchmark forecast. First estimation period: 1982:1-2000:1

Table 5: **RMSE from Full Sample with Additional Recursive Steps**

Horizon	**Forecast Method**				
Quarters	SVAR I(1)	SVAR I(0)	HP Filter	BP Filter	EXP Smooth
1	1.01*	1.05*	1.02*	1.06	1.10
2	1.03*	1.10	1.06*	1.17	1.24
3	1.06*	1.12	1.14	1.32	1.38
4	1.08*	1.19	1.26	1.46	1.52
5	1.12*	1.26	1.37	1.59	1.64
6	1.16*	1.31	1.47	1.69	1.73
7	1.22	1.32	1.54	1.73	1.75
8	1.25	1.33	1.58	1.78	1.78

	Forecast Method				
Quarters	Persistence	Ex. Food Engy	Trim	WTrim	Factor
1	1.03*	1.18	1.23	1.26	1.02*
2	1.06*	1.37	1.43	1.52	1.06*
3	1.08	1.53	1.55	1.71	1.12
4	1.09	1.63	1.59	1.84	1.20
5	1.08	1.73	1.60	1.93	1.29
6	1.05*	1.76	1.54	1.94	1.39
7	1.01*	1.72	1.46	1.88	1.44
8	1.00*	1.71	1.38	1.83	1.52

Note: The table shows the ratio of the RMSE from a regression with the named core inflation rate to the "no change" forecast. A value less than one signifies a lower forecast error than the benchmark forecast. First estimation period: 1960:1-1993:1.* indicates that the null hypothesis of equal predictive ability is rejected at the 10% level. Rejection of null indicates that core inflation forecasts are not statistically inferior to benchmark. In no case are the core forecasts statistically superior however.

Table 6: Directional Forecasts for Core Measures

	Long, r = 28		Short, r = 28		Long, r = 56	
Core Measure	Q = 4	Q = 8	Q = 4	Q = 8	Q = 4	Q = 8
HP Filter	0.32	0.50	0.25	0.29	0.37	0.41
Persist	0.21	0.46	0.50	0.39	0.41	0.45
SVAR I(1)	0.68	0.57	0.43	0.39	0.59	0.52
SVAR I(0)	0.32	0.43	0.64	0.43	0.39	0.32
Ex. Food Engy	0.32	0.50	0.32	0.54	0.38	0.48
BP Filter	0.11	0.36	0.14	0.36	0.20	0.34
EXP Smooth	0.21	0.32	0.29	0.39	0.38	0.45
WTRIM	0.43	0.46	0.43	0.50	0.45	0.43
TRIM	0.54	0.58	0.36	0.39	0.46	0.46
Factor	0.50	0.39	0.57	0.54	0.48	0.48

Note: The table shows the percentage of the time that the core inflation rate correctly predicts the direction of future price changes four quarters and eight quarters ahead. Like the previous forecast exercise, results are presented for the full sample with 28 recursive steps, the post Moderation sample with 28 recursive steps and the full sample with 56 recursive steps. One would expect a good forecast model to beat a coin flip in the sense that it would forecast the direction of inflation correctly 50% of the time.

Table 7: Concordance of Core Inflation Measures

Core Measure	Persist	SVAR I(1)	SVAR I(0)	HP Filter	Ex. Food Engy
Persist	1.00	0.63	0.59	0.55	0.64
SVAR I(1)	0.63	1.00	0.49	0.53	0.61
SVAR I(0)	0.59	0.49	1.00	0.54	0.58
HP Filter	0.55	0.53	0.54	1.00	0.79
Ex. Food Engy	0.64	0.61	0.58	0.79	1.00
BP Filter	0.63	0.68	0.56	0.70	0.79
Exp Smooth	0.58	0.58	0.51	0.88	0.89
WTrim	0.69	0.59	0.43	0.66	0.65
Trim	0.68	0.63	0.44	0.68	0.69
Factor	0.60	0.63	0.44	0.63	0.66
AVERAGE	0.62	0.59	0.51	0.66	0.70

Core Measure	BP Filter	Exp Smooth	WTrim	Trim	Factor
Persist	0.63	0.58	0.69	0.68	0.60
SVAR I(1)	0.68	0.58	0.59	0.63	0.63
SVAR I(0)	0.56	0.51	0.43	0.44	0.44
HP Filter	0.70	0.88	0.66	0.68	0.63
Ex. Food Engy	0.79	0.89	0.65	0.69	0.66
BP Filter	1.00	0.78	0.64	0.65	0.55
EXP Smooth	0.78	1.00	0.71	0.78	0.68
WTrim	0.64	0.71	1.00	0.84	0.79
Trim	0.65	0.78	0.84	1.00	0.85
Factor	0.55	0.68	0.79	0.85	1.00
AVERAGE	0.66	0.71	0.67	0.69	0.65

Note: The table indicates the degree to which different core measures agree on the sign of the inflation gap. The table needs to be read as a grid reference. For example, the number 0.63 in the second row of the first column indicates that the persistence measure and the SVAR I(1) agree on the sign of the inflation gap 63% of the time. Averages are also provided for each core measure.

Chapter 3

Quantifying the Impact of Oil Prices on Inflation

Core inflation rates are widely calculated. The perceived benefit of core inflation The substantial increase in oil prices over the past six or seven years (up to 2008) has provoked considerable comment within the international media. While this increase has not had quite the same impact as that experienced in the 1970's, the magnitude of the price increases still has significant implications from a macroeconomic perspective. This is particularly the case in terms of inflation. The re-emergence of the oil price issue necessitates a re-examination of econometric estimates of the influence of oil prices on inflation. We examine this issue in the case of a small open economy - that of Ireland.

3.1 Introduction

The observed reduction in the volatility of macroeconomic variables in developed economies, commonly referred to as the "Great Moderation", has been well documented. The salient features for these economies has been low, stable inflation and consistent economic growth. The stability of inflation is now under serious threat. Inflation has been on the rise in developed economies following persistent increases in the prices of oil and agricultural commodities. The impact of oil prices on inflation, in particular, has re-emerged as a key macroeconomic issue. During 2003, oil traded between \$28 and \$36 per barrel but by May 2008, this had increased to almost \$125 per barrel. This has had a significant effect on inflation internationally.

To quantify the impact of oil prices on inflation, consider figures released by Eurostat on euro area inflation. The Harmonised Index of Consumer Prices (HICP) is a pan-European measure of inflation. Amongst others, it is calculated for the individual countries of the EU and for the Euro Area as a whole. Eurostat also produce a core measure of the HICP which excludes energy prices. The difference between the headline rate and the core rate is the contribution of the energy component to the headline rate. The HICP inflation rate for the Euro Area in May 2008 was 3.7%. The inflation rate excluding energy was 2.6%. Thus, the contribution of the energy component was 1.1%, meaning that roughly one third of overall inflation was attributable to energy. Clearly, the energy component is one of the most important contributors to the overall inflation rate in the current climate. The impact of energy price developments on inflation is easy to quantify after the fact however. The key question this paper aims to address is whether it is possible to construct forecasts of the energy component based on current oil price developments.

To answer this question, we consider the case of Ireland. Ireland represents the archetypal case of a small open economy. In 2007, the sum of imports and exports was equivalent to 148% of GDP, meaning that, as a highly open economy, Irish inflation rates are significantly affected by international developments. The Irish inflation rate has been pushed higher by international oil and food prices in the same way as other European economies. In the European context, the value of the euro against the dollar has also increased significantly. In the Autumn of 2002, the euro and dollar were broadly equal in value. From March 2008 until July 2008, representing the last five months of data, one euro has rarely traded below one dollar and fifty five cents. The strength of the euro has insulated countries within the Euro Area from the full effects of the recent oil price spikes but these economies have still experienced considerable energy inflation.

The approach taken in this paper is purely empirical. The aim is to find a modelling approach that is optimal from a forecasting perspective. Forecasts are constructed for three months into the future, as model-based energy forecasts are extremely poor at even medium term horizons. However, in order to compute forecasts of the energy component over longer horizons, data on oil price futures are used to condition forecasts from the short-term forecast models. The energy component is broken down into its constituent parts and stripped of administered price series. Administered prices are the prices of items that are either fully or partially regulated. In the context of the Irish energy component, this refers to the electricity and piped gas series. By excluding these series, we have a more refined measure of energy inflation that is determined purely by market forces. This will be referred to as the Non-Administered Energy (NAE) series.

The constituent items in this series are forecast using two econometric methods. For each method, these individual forecasts are then combined to construct a forecast of the NAE series. The same forecasting methods are also applied directly to the NAE series. There are gains to be made from forecast aggregation but only over the shortest of forecast horizons. Using standard benchmark forecasts, it is possible to improve on the benchmark model for all forecast horizons. Reductions in the Root Mean Squared Error (RMSE) of the model forecasts relative to a benchmark range from 20% to 33% at the one-month horizon and from 9% to 26% at the three-month horizon. This represents a significant improvement in forecasting power. The paper also considers whether it is possible to improve forecast accuracy further by using the price of refined oil products rather than crude oil prices. Although there is not a universal improvement in forecast accuracy, there are meaningful improvements in certain cases. The next section describes recent developments in Irish inflation with a focus on the energy component. Section 3 reviews the literature and outlines the data used in the study. Section 4 contains an explanation of the empirical models used in the paper while section 5 presents the results of a number of forecasting exercises. Section 6 demonstrates how the model could be used for longer forecasting horizons while section 7 concludes the paper.

3.2 Inflation in an Irish Context

Energy price inflation has rarely been out of the business pages over the last three years. To put the oil price increase in context, Figure 1 graphs the price of a barrel of oil, which is priced internationally in US dollars, over the last eleven years. The more relevant measure of oil prices in Ireland is the euro price of oil so the graph also includes this series. It can be seen that the dollar price of oil, denoted in blue, has increased more rapidly than the euro price over the last five years as the strength of the euro has insulated those within the euro system from the full impact of the oil price increase. Despite the mitigating effect of these currency movements, there has still been considerable energy price inflation. Figure 2 graphs energy price inflation over the last five years. Over this period, year-on-year energy price inflation has generally been significantly higher and more volatile than overall inflation. Energy price inflation in May 2008 recorded an annual increase of 9.2% whereas the overall HICP rate was 3.7%. The annual average year-on-year change in the energy component was

12.6% in 2005, which was the largest recent annual change.

The high rates of energy inflation are reflected in the contribution of the energy component to overall inflation. Figure 3 graphs this contribution in percentage terms over the past ten years. There are periods when the contribution of the energy component has exceed 50%, meaning that if overall inflation was 2%, over 1% of this would be driven by the energy component. Over 2005, the influence of energy prices was particularly strong with a contribution of nearly 40% on average over the year. Given the importance of the energy component of inflation, it is critical to have a clear understanding of the impact of oil price changes on energy inflation.

The weight of the energy component in the HICP is approximately 8.7% for Ireland. This is slightly higher than its weight of 7.8% in the CPI as the HICP is a smaller basket of goods and services. The composition of the energy components in the HICP and CPI are identical however. Thus, although the focus of this paper is the HICP, the results are equally valid with respect to the CPI energy component. In this paper, the energy component is split into its constituent parts. The first column in Table 1 shows the current weights of the various elements that constitute the energy component. The exact weight of each item and the mix of items in the component changes every five years when the CPI is rebased but changes to the make-up of the energy component over the last fifteen years have been fairly minor.

Unleaded petrol and diesel together account for approximately 48% of the energy component, but with petrol having a weight almost four times that of diesel. Home heating oil is referred to as fuel oil in this paper - this is how it is referred to in the HICP basket by our statistical agency. It accounts for 11% of the index. These three items are all heavily influenced by oil price developments. Other items generally used for home heating such as coal, turf, briquettes and fire-lighters together account for under 10% of the energy component. Electricity is an important item, with a weight of 20%. Of the two gas components, piped gas, at about 10% of the index, is much more important than bottled gas, which only has a weight of 1.5%. However, piped gas and electricity prices are both regulated and change price only occasionally and in discreet jumps. As they do not respond in a predictable way to international energy prices, they were excluded from the current energy component in order to form a new market driven component. This removed approximately 30% of the weight of the official energy component. The second column of Table 1 shows the weights of the NAE series once the remaining items are rescaled following the removal of the administered price series.

3.3 Modelling and Forecasting Inflation

3.3.1 Review of Literature

The paper is primarily a forecasting paper, which estimates the impact of oil prices on inflation. There is an existing literature that aims to quantify the impact of oil prices on key macroeconomic variables such as output and inflation. Amongst others, Hamilton (2008) notes that nine out of ten US recessions since World War II were preceded by a spike in oil prices. A number of studies discount the possibility that the relationship between oil prices and output is merely a statistical coincidence.[1] In the past surging oil prices have stalled growth and employment but stimulated inflation, the combination of which is referred to as stagflation. Thus, the potentially ruinous effects of rampant oil prices are already well documented.

One of the most active research areas of late is in attempting to explain why current oil price increases have not had the same impact on the world economy or national economies as the oil price shocks of the 1970's. Gregorio et al (2007) highlight a number of potential reasons including decreased oil intensity in production and a reduction in exchange rate pass-through. Chen (2008) suggests that a low inflation environment may also contribute to a reduced impact of oil prices on inflation. This change in the reaction of national economies to recent oil price shocks relative to earlier shocks suggests that econometric estimates gauging the impact of oil need to be revised and updated. This paper estimates the inflationary impact of oil prices in the Irish context. The impact for Ireland reflects that for other small open economies although the precise quantitative estimates will vary according to the dependency of each country on its oil imports. Standard econometric models are used in the analysis but the approach is non-standard in a couple of respects. We now outline some of the paper's contributions.

Models that aim to predict the inflationary impact of oil prices typically specify a model in terms of the aggregate inflation rate. Papers such as Chen (2008) and Gregorio et al (2007) do this in a panel set-up while others such as Hooker (2002) focus on the US response. However, there are always other factors at work in the economy besides oil prices that are impacting on the aggregate inflation rate. For example, the significant increase in agricultural commodity prices through 2007 and 2008 could, if not properly controlled for, be mistaken for oil price inflation. The approach in this paper is to examine the impact of oil prices on the energy component in isolation. This will control for other factors, as items such as food price increases will not impact on a narrow price index measuring only energy changes. This is not the first paper to look at specific energy price indicies. Chacra (2002) examines the impact of oil price changes on the specific elements of the energy component of the CPI in Canada with a view to forecasting the individual components but does not consider forecasts for the aggregate component. Following Chacra (2002), individual components of the

[1] See Hamilton (1983), Rotemberg and Woodford (1996) and Carruth et al (1998) for example.

energy index are also forecast in this paper. However, these forecasts are combined and then compared with direct forecasts of the aggregate.

Regulated prices are cleansed from the energy component so that a purely market driven measure of energy inflation is the focus of the paper. Oil prices will have greater predictive ability for a series of this type. Another issue considered in this paper is which oil price series should be used for the models. The papers outlined above are based on prices for crude oil. Thus, they examine the pass-through or forecast power from crude oil to retail prices. This paper also considers whether it is better to use the price of refined petroleum products as these are closer in the distribution chain to the retail price and may improve forecasts. Borenstein et al (1997) also use both crude and refined prices but they focus on the issue of asymmetric pass-through but they are only concerned with gasoline prices. This paper considers the use of refined prices in broad forecasting context.

Accurate medium term forecasts of inflation are notoriously difficult to generate. This is reinforced by the vast array of models that fail to beat standard benchmarks at even moderate time horizons. Atkeson and Ohanian (2001) found that Philips curve models could not consistently beat a naive benchmark forecast of overall inflation. Stock and Watson (2006) point out that "the improvement of standard multivariate forecasting models, such as the backward-looking Philips curve, over a univariate benchmark has been less in percentage terms since the mid-1980s than before". Thus, the difficulty in beating standard benchmark inflation forecasts using multivariate models has increased. This motivates the choice of the Atkeson-Ohanian benchmark. However, as this benchmark is mainly used in relation to aggregate CPI inflation, which tends to be more persistent and less volatile than the inflation rate of the energy component, this paper also considers a second benchmark in the form of a autoregressive forecast.

The approach taken to forecasting Irish inflation has been quite varied in terms of the technical tools used but a key unifying theme in the literature is the recognition that the open nature of the economy is a critical factor in the determination of Irish prices. Slevin (2001) notes that the output gap, which is a purely domestic measure, is not sufficient to explain Irish inflation. Kenny and McGettigan (1996) and Slevin (2003) both use small open economy models, in which there is a distinction between the traded and non-traded sectors, in explaining Irish inflation. Kenny and McGettigan (1996) also model imports prices specifically using an exchange rate pass-through model. Bermingham (2007) calculates a core inflation measure taking account of the role of oil prices and uses the core measure to forecast HICP inflation. This paper, however, is the first to focus explicitly on the energy component of inflation in a forecasting context.

3.3.2 Data

The data used in the study are monthly. Oil prices refer to the price per barrel of crude oil denominated in US dollars and are available from the IMF. These were converted to euro using average monthly exchange rates. The price series for the constituent energy components were constructed by chain-linking these series across three CPI base periods. The NAE series was then constructed using these series. These series are available from December 1996 - May 2008, which results in a sample of 138 data points prior to any variable transformations. The refined prices used in the paper are the Rotterdam gasoline and diesel prices. These series are available daily but monthly averages were again taken for this paper.

3.4 Empirical Approach

Two forecasting models are used in the paper. These forecasting methods are applied to the aggregate NAE energy component. Solely quantifying the effect of oil prices on this measure of inflation only is open to the criticism that second round effects are ignored but there is limited evidence of second round effects to date. Second round effects refer to the situation where oil price increases lead to general inflation in other sectors of the economy through increased production costs. Some second round effects have been seen in the transport sector, particularly in the airline industry. There has also been an increase in some regulated prices but these cannot be modelled econometrically. Outside the energy and food sectors, inflation in other sectors is not running significantly above historical levels, suggesting limited evidence of second round effects.

The forecasting methods are also applied to the individual items in the component and the forecasts are aggregated to arrive at a second candidate forecast for the energy component. This aggregation procedure is carried out for both forecasting models. It is found that there are gains to computing individual forecasts and then aggregating the forecasts. However, these are greatest at the one-month time horizon. Of the eight items in the NAE series, only three are forecast individually and naive forecasts are used for the remaining five. Models are used to forecast petrol, diesel and fuel oil. The only data used to forecast these series are the series themselves and oil prices. Naive forecasts are made for firelighters, coal, turf and briquettes, motor oil and bottled gas. The items which are forecast using the naive approach account for only 16% of the NAE series. Some items are not forecast because of their small weight in the index and others because they have no relationship with oil prices. Direct forecasts are also constructed using the prices of refined oil products as the inputs to the estimation and forecast procedure.

3.4.1 Forecast Benchmark

The first benchmark forecast is taken from Atkeson and Ohanian (2001) and will be referred to as the AO benchmark in the remainder of the paper. The forecast horizon is three months. The benchmark forecast is for no change in the current inflation rate. For example, if yearly energy inflation is 4.5% for June 2005, the benchmark forecast for each month for July 2005 - September 2005 is also 4.5%. It is a naive forecast but the literature mentioned previously confirms that it can be difficult to beat in many instances. The second benchmark considered is a simple autoregressive model. Forecasts are performed on a recursive basis. The first sample period for recursive estimates is December 1996 - March 2005. Models are estimated over this time frame and forecasts for April 2005 - June 2005 are computed. One month is added to the sample and the process is repeated. In this way, 36 sets of forecasts at the three month horizon were computed for each method. Forecasts are evaluated using the Root Mean Square Error (RMSE), a standard approach in the forecasting literature.

3.4.2 Modelling Strategy

The strategy used to model the NAE inflation series and its constituent parts is based on the strong observed co-movements between these price series and oil prices. Figure 4 depicts oil price inflation and inflation of the NAE component. The similarity in the behaviour of the two series is striking, with movements in oil price inflation tending to lead those in energy inflation with a small lag. On further inspection, it is clear that the individual items in the energy component display the same patterns. Figure 5 graphs oil price inflation and inflation in petrol, diesel and fuel oil inflation. The final part of the graph shows bottled gas inflation and international gas price inflation. Petrol, diesel and fuel oil prices are clearly driven by international oil price developments. Again, oil prices seem to demonstrate a leading relationship. Having uncovered a strong relationship between oil prices and energy inflation rates at both an aggregate and disaggregate level, the obvious question is whether it is optimal to forecast the individual components or instead the aggregate series. We consider both alternatives.

Prior to the discussion of the models used, let us first establish the stochastic properties of the data. As is typical, all the price series are non-stationary - results of unit root tests are not presented for this in the interests of brevity but are available upon request from the author. The inflation rates of oil and the energy components are all stationary. The results of the unit root tests are presented in Table 2. The tests are for the year-on-year growth rates. The unit root test used was the standard Augmented Dickey-Fuller (ADF) test. The critical value for this test at the 5% level given the sample size is -2.88. For all variables, the test statistic comfortably exceeds the critical value and so the null of a unit root is rejected for all the inflation rates considered.

Moving to the estimation strategy, the first model used is a standard pass-through equation in the form of an Autoregressive Distributed Lag (ARDL) model. In this set-up, the inflation rate of a certain item is regressed on past values of itself and past oil price inflation:

$$\pi_t = \alpha_0 + \sum_{i=1}^{n} \beta_i \pi_{t-i} + \sum_{i=1}^{n} \theta_i oil_{t-i} + \epsilon_t \tag{3.4.1}$$

This type of equation was fitted to the NAE series and to its components. Generally speaking, it was found that two lags for each variable fit the data and produced the best forecasts although there is a slight variation for some series. The residuals from the estimated equations were tested for serial correlation. The presence of lagged dependent variables in the regression creates a bias towards a finding of no serial correlation with the Durbin-Watson statistic. The general LM test suggested by Godfrey (1978) and Breush (1978) is used instead. There is no evidence of serial correlation using this test for either the NAE series or any of its components.[2]

In order to generate the forecasts, past values of oil price inflation and the past inflation rate of the series are needed. At the one month horizon, the forecasts can be conditioned on actual observed data. Beyond this, earlier period forecasts of the inflation rate can be used to condition later forecasts. The following equations detail how the three forecasts are constructed in each recursive step, assuming two lags for both energy inflation and oil price inflation. A "hat" is used to denote a forecasted variable:

$$\hat{\pi}_{t+1} = \alpha_0 + \beta_1 \pi_t + \beta_2 \pi_{t-1} + \theta_1 oil_t + \theta_2 oil_{t-1} \tag{3.4.2}$$
$$\hat{\pi}_{t+2} = \alpha_0 + \beta_1 \hat{\pi}_{t+1} + \beta_2 \pi_t + \theta_1 oil_{t+1} + \theta_2 oil_t \tag{3.4.3}$$
$$\hat{\pi}_{t+3} = \alpha_0 + \beta_1 \hat{\pi}_{t+2} + \beta_2 \hat{\pi}_{t+1} + \theta_1 oil_{t+2} + \theta_2 oil_t \tag{3.4.4}$$

In the first forecast, actual data can be used for the two lags of both energy price and oil price inflation - there are no hats on the right hand side of Equation 2. In period $t+2$, the first lag needed to construct the forecast is from period $t+1$. The forecast of inflation constructed in the first forecast step is used as the first lag for energy price inflation in the second forecast. The second lag still refers to actual data. Thus, in the second forecast equation, we see that the first lag of inflation has a hat while the second does not. For oil prices, forecasts are constructed using the last available data points even though these don't technically represent the lag from period $t+1$. This approach is taken because it yields the most accurate forecasts. For Equation 4, the energy price inflation lags are themselves both forecasts, as indicated by the two hats, while oil inflation lags are again the last actual data points.

Having already established the stochastic properties of the price series, they were tested for cointegration with oil prices using the Engle-Granger method. The results

[2]Results not reported but available upon request.

are presented in Table 3. In each case, the test was based on a cointegrating vector with the named variable and international oil prices converted to euro. The variables all exhibit evidence of cointegration at the 5% level. Given this evidence, long-run equations were estimated for each pair of variables. Short-run equations including an error correction term (ECM) were then estimated and used to construct the forecasts. In contrast to the ARDL approach, some assumptions need to be made about the future path of oil when constructing the forecasts from this approach, as future values of the equilibrium error are needed. Two assumptions were tested - oil prices remain constant or oil inflation remains constant. The constant price assumption was found to generate slightly better forecasts so this is the assumption used. The constant price assumption, together with the forecast of the retail price, allows the equilibrium error to be calculated for future periods. This equilibrium error is then augmented to the equations above.

3.5 Forecasts

3.5.1 Forecast Evaluation

The forecast methodologies described in the previous section are now evaluated. Table 4 presents the results of the different forecasting methods applied to the individual components and to the NAE series directly. The numbers in the table the RMSEs and each section of the table shows the errors for a specific component. In each section, the first two rows show the errors from the two benchmarks - the Atkeson-Ohanian forecast and the autoregressive forecast. The third and fourth rows of each section show the errors for the ARDL forecast and the cointegration model forecast.

The first section presents the results for the NAE series, the market driven energy series constructed in the paper. The ARDL forecasts and the cointegration forecasts are more accurate than both benchmarks. Although the ARDL and cointegration models have similar forecast power, the cointegration forecasts are marginally more accurate at all forecast horizons. The improvements in forecasts power relative to the AO benchmark using the cointegration approach are 21%, 19% and 21% at the one, two and three-month forecast horizons respectively.

A similar picture emerges in the remaining sections of the table, which detail forecast performance for the three main components of the NAE series. For each component, the two econometric forecasts outperform the benchmark forecasts. In addition, the forecasts using the cointegration model are slightly more accurate than those of the ARDL model. Improvements in forecast power relative to the benchmark are greater at the shorter horizons. Using the cointegration approach, the one-month forecast of fuel oil is 33% more accurate than the AO benchmark whereas at the three-month horizon, the greatest improvement in forecast power is for the petrol forecast, which is 26% more accurate than the benchmark.

3.5.2 Assessing the Importance of Oil Prices

The importance of oil prices can be gauged by estimating a purely autoregressive model and the ARDL model. The only difference between these two models is the inclusion of oil price inflation in the ARDL model. If oil prices are really helping to improve forecast performance, we would expect forecasts without the oil prices included to perform poorly. The results of this exercise are presented in Table 5. The table presents the ratio of the RMSE from a model including oil prices to a purely autoregressive model for the NAE series and for individual components. At all time horizons, this ratio is less than one indicating that oil prices are contributing to forecast performance. For the direct forecasts of the NAE series, forecasts are 18% more accurate at the one-month horizon and 14% more accurate at the two and three-month horizons when oil prices are included. As a general feature, the improvement in forecast accuracy is more pronounced at the shorter horizons given the short lag lengths in the model.

It is also possible to test more formally if the ARDL model produces forecasts which are statistically superior to those of the AR model which excludes oil prices. For this exercise, only direct forecasts of the NAE series are compared. If the model with oil prices has more accurate forecasts statistically, this is equivalent to saying that the difference between the two forecast errors is statistically significant. Given that we are only interested in an improvement in forecast power relative to the autoregressive model, the hypothesis test is one-sided. The mean squared error (MSE) is used in place of the RMSE in the test. As is suggested from their names, the RMSE used up to now is simply the square root of the MSE. The MSE from the two models is compared using a statistic that identically resembles a standard t-test. Under the null of equal predictive ability, the form of the statistic is given by:

$$H_0: \ \delta_1^2 - \delta_2^2 = 0; \ \ S = \frac{\hat{\delta}_1^2 - \hat{\delta}_2^2}{\left(\hat{V}/P\right)^{\frac{1}{2}}} \tag{3.5.5}$$

where δ_1^2 is the MSE from the null (autoregressive) model, δ_2^2 is the MSE from the alternative survey model, $\hat{V}$ is the estimated variance of the forecast differential series and P is the number of predictions or forecasts. Despite the familiar form of the test, there are two potential complications when calculating this statistic.

One possible complication arises from the fact that the series of forecast error differentials used to construct the statistic can be serially correlated. This is normally the case when forecasts are performed for horizons beyond one-step. The reason for this is that forecast periods overlap for multistep forecasts in consecutive recursive iterations. When serial correlation is present, the long-run variance needs to be estimated. Correcting for serial correlation by using the long-run variance and then using standard critical values is referred to as the Diebold-Mariano (1995) test. The long-run variance is calculated as the spectral density of the forecast differential series at frequency zero. The Newey-West non-parametric kernel estimator is used with the automatic bandwidth selection procedure suggested by Andrews (1991). The difference between the standard variance and the long-run variance is small in this application.

A second complication arises when the null model is a nested version of the alternative model. In this case, it can be seen that the autoregressive model is nested in the ARDL model - the ARDL model reduces to the autoregressive model when the coefficients on oil prices are restricted to zero. Assume the null to be true, so that the autoregressive model is the true model. In the ARDL model, there are additional estimated regression parameters. The values of these coefficients are zero in population. They will not be exactly zero in-sample due to parameter estimation error. When performing out-of-sample forecasts, the additional noise imparted in the forecasts from including parameters whose population values are zero means that the mean squared prediction error will be larger for the alternative model i.e. $\delta_2^2 > \delta_1^2$ so that the hypothesised difference in MSE, $\delta_1^2 - \delta_2^2 < 0$. This means that the test statistic is not centred at zero - it is centred in negative territory. Standard test statistics are based on distributions with a mean of zero.

There are a number of ways to correct for this problem. The correction used in this application, which is the easiest computationally, is based on Clark and West (2006 & 2007). They recommend re-centring the distribution at zero, using a correction based on the fitted values (forecasts). The specific correction depends on the form of the null model. Given the parameterisation of the null in this application, the adjustment term equals the mean squared forecast differential. This adjustment term is added to the numerator of the test statistic displayed above. Having carried out this adjustment, inference can proceed in the usual fashion using conventional, asymptotically normal procedures familiar from Diebold and Mariano. The results of this exercise are presented in Table 6. The table shows the test statistics under the null that the AR model and ARDL model have equal predictive ability. Statistics are presented for each forecast horizon and the 5% critical value is 1.645 in each case. The null is rejected at all time horizons indicating that the inclusion of oil prices to the basic AR model results in statistically significant improvements in forecast performance. The forecast errors from the AO benchmark are larger than those from the AR model. Thus, we can be relatively certain that the improvement in forecast accuracy relative to the AO benchmark is also statistically significant.

3.5.3 Forecast Aggregation

Turning to the issue of forecast aggregation, the individual forecasts from the two estimation techniques were combined with naive forecasts for the components that are not modelled explicitly to form a second forecast for the NAE series. The intuitive approach to combining the forecasts would be to take the weights in Table 1 and multiply them by the forecasted inflation rate for each item. Consider this calculation in the simple case in which an aggregate series x at time t is made up of just two different series, x_1 and x_2 with weights λ and γ respectively that are fixed from the base period. This represents the case of a Laspeyers price index such as the HICP in Ireland:

$$x_t = \lambda x_{1,t} + \gamma x_{2,t} \tag{3.5.6}$$

The inflation rate, π_{t+h}, of the aggregate between t and $t+h$ is defined as its percentage change:

$$\begin{aligned}
\pi_{t+h} &= \frac{x_{t+h} - x_t}{x_t} \\
&= \frac{(\lambda x_{1,t+h} + \gamma x_{2,t+h}) - (\lambda x_{1,t} + \gamma x_{2,t})}{\lambda x_{1,t} + \gamma x_{2,t}} \\
&= \lambda \frac{x_{1,t+h} - x_{1,t}}{\lambda x_{1,t} + \gamma x_{2,t}} + \gamma \frac{x_{2,t+h} - x_{2,t}}{\lambda x_{1,t} + \gamma x_{2,t}}
\end{aligned} \tag{3.5.7}$$

The weighted average of the inflation rates of the two individual series is given by this expression:

$$\lambda \pi_{1,t+h} + \gamma \pi_{2,t+h} = \lambda \frac{x_{1,t+h} - x_{1,t}}{x_{1,t}} + \gamma \frac{x_{2,t+h} - x_{2,t}}{x_{2,t}} \tag{3.5.8}$$

Clearly, the two expressions are not equal. Thus, with a fixed weight price index, it is not the case that the weighted average of the inflation rates of the individual items is equal to the inflation rate of the overall index. This lack of additivity only relates to the change in the index and not to the level. In other words, the weighted average of the price level of the individual items does equal the price level of the aggregate. In order to aggregate, we use the forecasted inflation rates to generate a forecast of the weighted price level for each item, sum the weighted prices and calculate the implied inflation rate for the aggregate.

Table 7 presents the results of this exercise. The numbers in the table refer to the ratio of the RMSE from the disaggregate approach versus the aggregate approach. A value less than one again indicates that the combination of individual forecasts is more accurate than using the same approach to forecast the NAE series directly. The results for the ARDL approach show that forecast combination can improve forecasts at the one-month horizon by 6% relative to a direct forecast. At months two and three, the forecast errors are broadly similar using indirect versus direct forecasts. For the cointegration model, there is a 8% improvement at the one-month horizon, no difference at two months and a 3% improvement at three months. Overall, the results suggest some role for forecast aggregation, as the results are encouraging at the one-month horizon and, although the gains may be modest at other horizons, it is rarely the case that the combined forecasts are less accurate.

3.5.4 Refined Prices

In this section, we consider an alternative data source for the input to the forecasts. Specifically, data on the refined price of oil products is used in place of international crude oil. The difference between the two can be thought of as the cost of refining crude oil into a product suitable for retail distribution. Refined prices for gasoline and diesel are available and these represent the prices that refineries charge retailers for gasoline and diesel. Although the price paid at the pump, and in turn reflected in the consumer price index, will also incorporate the profit margin of the retailer and any local taxes, the refined price is closer to the retail price than the price of international crude oil and may help in the construction of more accurate forecasts.

Direct forecasts of the NAE series and its components are constructed in the same manner as before. The refined price of gasoline is used as the input in the NAE and the petrol price forecasts while the refined price of diesel is used to construct the diesel and fuel oil forecasts. The results are presented in Table 8 and are analogous to those in Table 4. The first two rows of each section present the RMSE values for the benchmarks. These are identical to the numbers presented in Table 4 but are replicated here for convenience. As was the case with the forecasts based on crude oil prices, the refined price forecasts are more accurate than the two benchmarks considered at all time horizons.

If we compare the ARDL forecasts using refined prices in Table 8 to those using crude prices from Table 4, there are considerable improvements in forecast power at the one-month horizon for all items. The results are mixed at the two-month horizon. The forecasts for the NAE series and diesel are more accurate but petrol and fuel oil are less accurate. At the three-month horizon, only the forecast for diesel is more accurate. These results would appear to suggest that the benefits to using refined prices are confined to the short forecast horizons. However, for the NAE series, the improvements in forecast power at months one and two are quite large whereas the forecast at month three is only marginally less accurate, so the evidence in favour of using refined prices in the ARDL forecast of the aggregate NAE series is quite compelling.

For the cointegration model forecasts, there are some similarities in the results. At month one, all forecasts are more accurate when using refined prices relative to crude prices. At month two, the forecasts for the NAE series and for fuel oil are more accurate but petrol and diesel are less accurate while all forecasts using refined prices are less accurate for the three-month forecast. In Table 4, we saw that forecasts from the cointegration approach were generally more accurate than those from the ARDL approach. The reverse tends to be true with refined prices. Comparing direct forecasts of the NAE series using both econometric methods and both data types, one would favour the ARDL forecast with refined prices. Of all four methods, it has the most accurate forecasts for the first two months. The cointegration forecast with crude prices fares slightly better at month three but the ARDL with refined prices still has the best overall performance. The results suggest that forecast aggregation is unlikely to improve upon direct forecasts as the component forecasts are less accurate than the direct forecasts in the majority of cases.

3.6 Long-Term Forecasts

Despite the focus on short-term forecasts, in this section we generate long-term forecasts over a one-year horizon. Forecasting oil prices is now even more difficult than usual given their current volatility. The cause of the current oil price spike is difficult to attribute to any one cause. Many media commentators are pointing to the role of speculators as the driving force but this is disputed by others who maintain that fundamentals are driving the market. In this paper, the future path of oil is first taken from futures markets. The forecast is subject to a large degree of uncertainty which is the reason that short-term forecasts were favoured but the sensitivity of the forecasts to the oil price profile is examined by considering a second oil price path.

The data used in the estimation of the models ends in May 2008. Forecasts are constructed for June 2008-May 2009. A forecast of the exchange rate is also needed for the following year to construct forecasts and it is assumed that the euro/dollar rate remains unchanged. Actual oil price data are available for June 2008 but price data are not. June oil prices were noticeably higher than May and futures market data reflect this. The futures market profile for oil prices over the next year suggest that oil prices will stay over $140 per barrel. This price level is considerably higher than the recent data used in estimation. Oil broke $100 per barrel in February 2008 and had almost climbed to $125 in May. With futures suggesting over $140 per barrel for most of the forecast horizon, year-on-year forecasts conditioned on futures data suggest a sharp increase in energy price inflation for the next few months given that dollar oil prices are expected to be roughly twice the price they were in corresponding months last year.

Figure 6 graphs the forecast of the NAE series over the next year conditioned on oil remaining over $140 dollars per barrel. This oil profile suggests NAE inflation will peak at just under 17% in July before falling back to 14% in October and further back to under 9% in May 2009, the end of the forecast horizon. The NAE series represents approximately 6.1% of the HICP so at the forecasted peak of energy inflation in October, the NAE component will add just over 1% to the HICP inflation rate.

To examine the sensitivity of the forecast to the oil price assumption, a second more benign oil price assumption is considered. For this forecast, oil prices are assumed to fall slowly back to $100 per barrel by December 2008 and remain at that level for the remainder of the forecast horizon. This oil price profile is chosen arbitrarily for the sensitivity analysis. Figure 7 graphs the forecast of the NAE series in this case. The forecast of the series is quite similar to the previous forecast for the first six months of the horizon but NAE inflation falls much more rapidly in the second half of the forecast horizon. Indeed, under this assumption, NAE inflation is just 2.4% at the end of the horizon.

3.7 Summary and Conclusions

This study provides a means of quantifying the impact of oil price increases on inflation. The exercise is conducted within the context of a small open economy. Overall inflation rates, on an international basis, have been subject to two major influences over the past few years, that of oil prices changes and agricultural commodity price increases. To control for the impacts of agricultural commodity price increases on inflation, the approach adopted here is to focus on energy inflation and, in particular, a measure of energy inflation which is purely market driven. However, once the impact of oil prices on this component is known, it is trivial to calculate the impact on overall inflation.

In focusing on the energy component, this paper shows that simple econometric techniques significantly outperform standard benchmarks up to three months into the future. By forecasting the constituent parts of the energy series, the issue of forecast aggregation is also considered but gains in forecast accuracy are limited to the one-month forecast horizon. Beyond that, it is optimal to simply forecast the energy series directly. The paper also investigates whether the use of the price of refined oil products in place of the price of crude oil can improve forecasts. The results indicate that considerable improvements can be made at short time horizons, particularly in the case of the direct ARDL forecast of the energy series. The paper also constructs long term energy inflation forecasts for the next year and, as one would expect, these are quite sensitive to the assumed future path of oil prices.

There are a number of potential avenues for future work. At present, the forecast aggregation procedure only leads to benefits at the one-month horizon. Forecast models could be developed for the items which are presently forecast using naive methods. In addition, local taxes constitute a large percentage of the retail price of petrol and diesel in Ireland. By taking explicit account of this in the model set-up, further improvements in forecast accuracy may be possible. A further consideration is that the retail price of petroleum products may respond asymmetrically to price increases and price decreases. A model that takes allows for this could yield further dividends in terms of forecast performance although a longer time series of data may be necessary to consider this issue.

3.8 Figures and Tables

Figure 1: Oil Prices Denominated in Dollars and in Euro

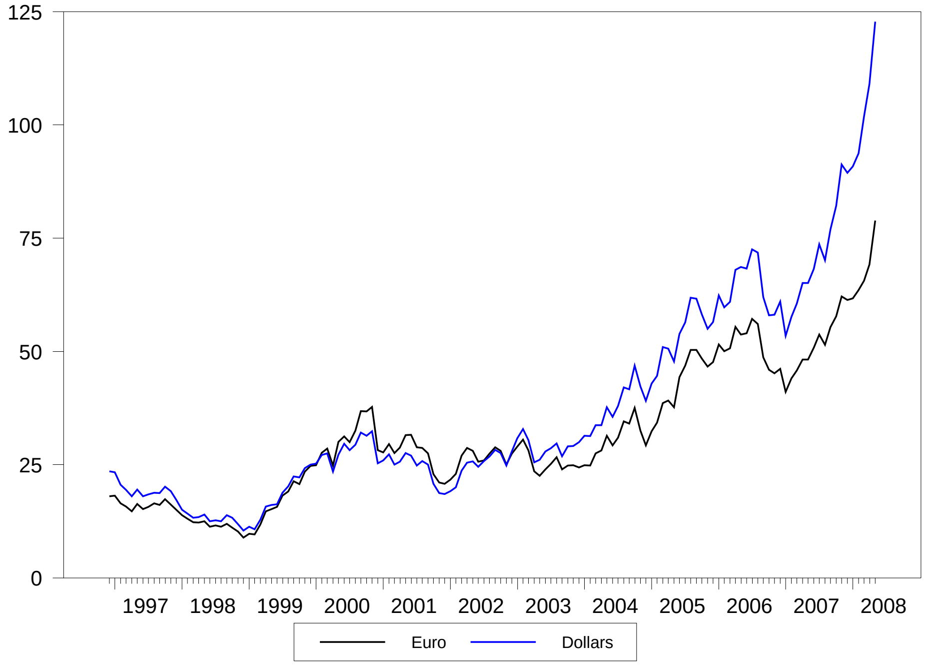


Figure 2: Recent Energy Price Inflation

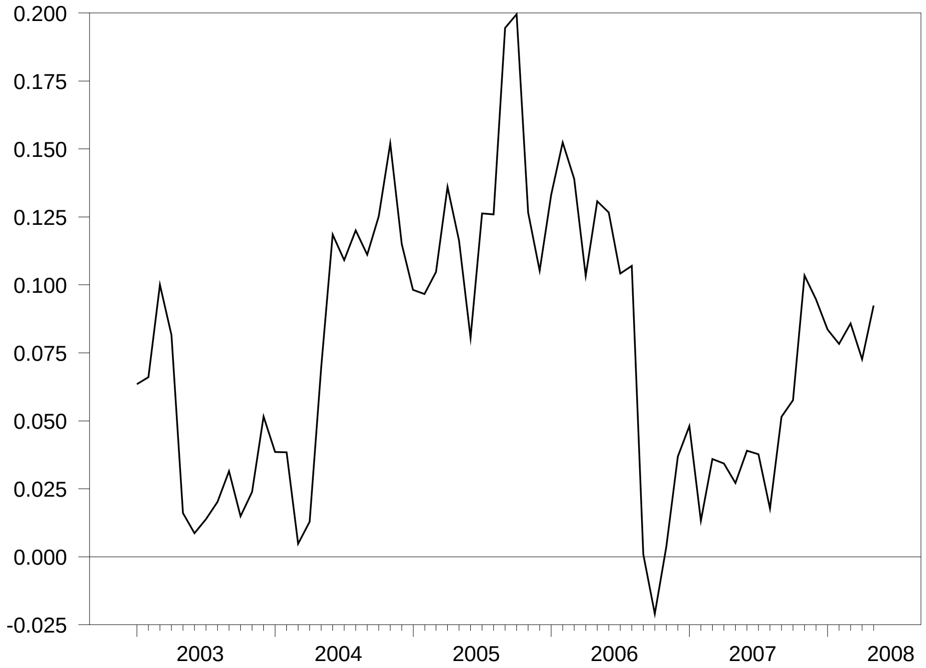


Figure 3: Contribution of Energy Component

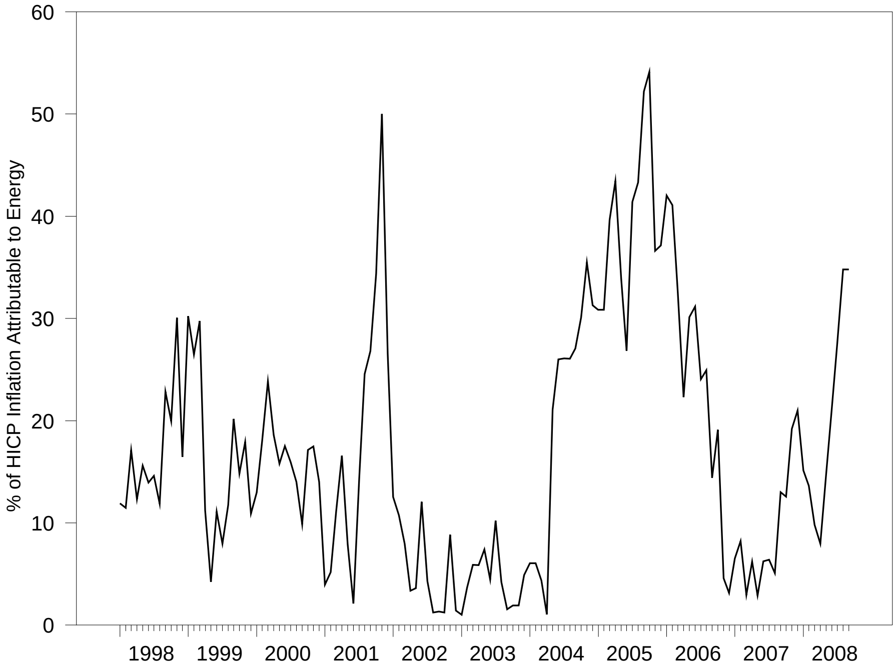


Figure 4: Oil Price Inflation and NAE Inflation

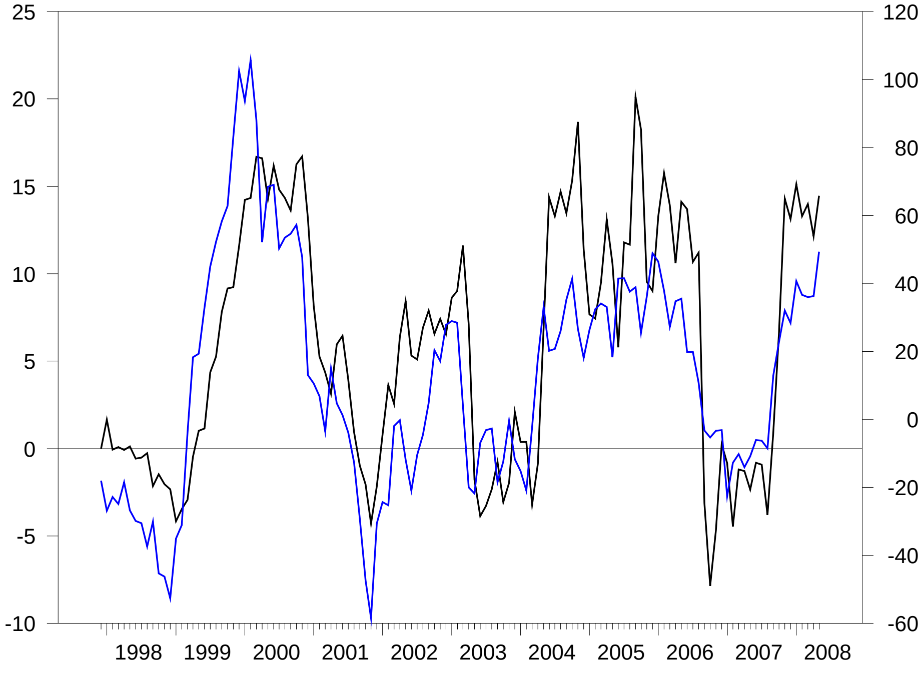


Figure 5: Energy Item Inflation and Commodity Prices

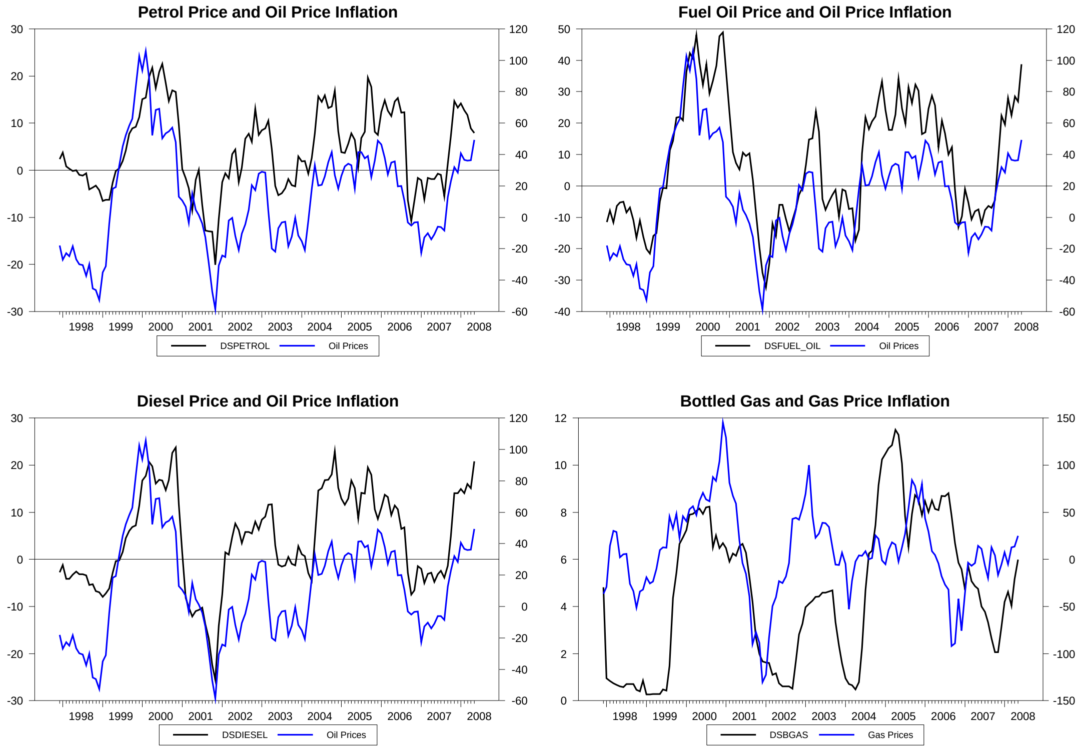


Figure 6: Forecast of NAE Series Inflation

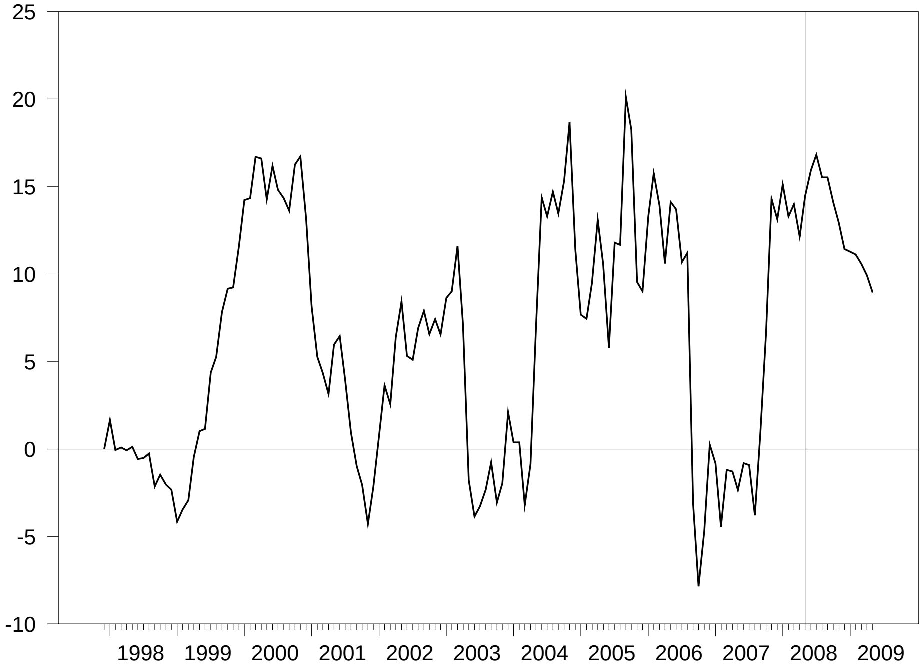


Figure 7: Alternative Forecast of NAE Series Inflation

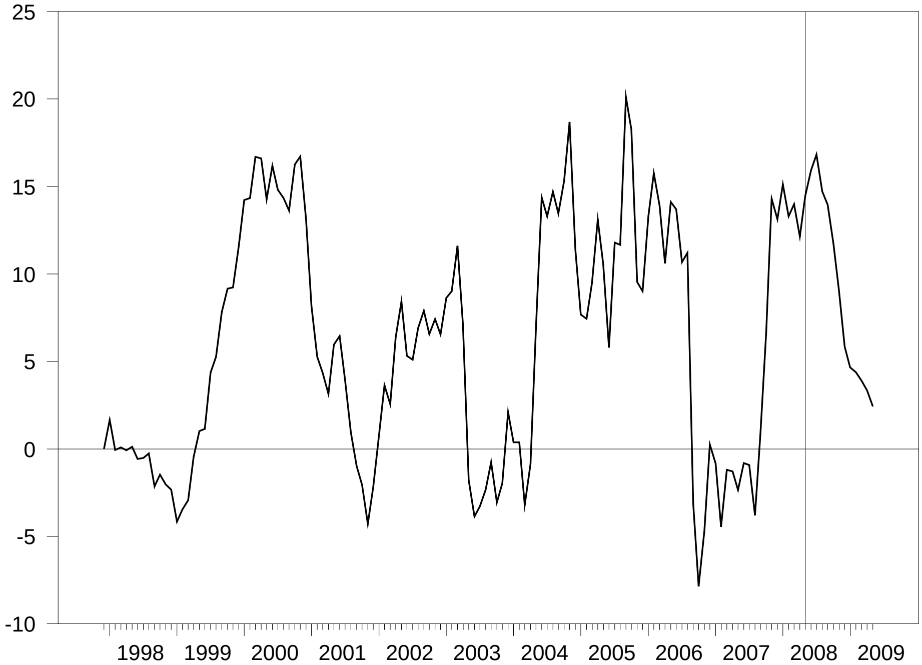


Table 1: Breakdown of the Energy Series

Item	Energy Weights	NAE Weights
Firelighters	0.63%	0.89%
Fire Handy Packs	0.12%	0.17%
Bottled Gas	1.52%	2.15%
Coal	4.42 %	6.28%
Piped Gas	9.74%	na
Electricity	19.89%	na
Fuel Oil	11.00%	15.63%
Turf and Briquettes	4.39%	6.24%
Petrol Unleaded	38.33%	54.46%
Diesel	9.79%	13.92%
Motor Oil	0.18%	0.26%

Table 2: Unit Root Tests

Variable	Statistic
Oil	-3.192^{+}
NAE	-3.3106^{+}
Petrol	-3.1324^{+}
Diesel	-3.0603^{+}
Fuel Oil	-3.4207^{+}

Note: The table tests the year-on-year growth rates of the variables in the Table for a unit root. + denotes rejection of the null hypothesis at the 5% level.

Table 3: Engle-Granger Cointegration Test

Variable	Statistic
NAE	-3.41^{+}
DIESEL	-4.57^{+}
PETROL	-5.11^{+}
FUEL OIL	-7.09^{+}

Note: The table presents results of tests to see if the named variables are cointegrated with international oil prices. Only two variables in each of the cointegration vector. Null hypothesis is that variables are not cointegrated. + indicates rejection of null at 5% level.

Table 4: Direct Monthly Forecasts of Energy Items

NAE	Month 1	Month 2	Month 3
AO	4.44	6.52	7.57
AR	4.33	6.29	7.12
ARDL	3.58	5.40	6.14
COINT	3.53	5.32	5.95
PETROL			
AO	5.20	7.75	9.09
AR	4.96	7.22	8.19
ARDL	4.10	6.42	7.16
COINT	4.05	6.25	6.75
DIESEL			
AO	3.83	5.77	7.02
AR	3.80	5.75	6.94
ARDL	2.89	4.74	6.34
COINT	2.90	4.76	6.28
FUEL OIL			
AO	7.55	10.13	11.79
AR	7.43	9.76	11.27
ARDL	5.40	7.98	10.73
COINT	5.03	7.69	9.81

Note: The table presents the RMSE for the benchmarks and the specified forecasting approaches.

Table 5: Contribution of Oil Prices to Forecast Accuracy

Method	Month 1	Month 2	Month 3
NAE	0.82	0.86	0.86
Petrol	0.83	0.89	0.70
Diesel	0.76	0.82	0.91
Fuel Oil	0.73	0.82	0.95

Note: The table presents the ratio of the RMSEs from the ARDL model to forecasts from a purely autoregressive model without lags of prices included. A value less than one indicates that the model with oil prices included is more accurate.

Table 6: Tests of Equal Predictive Ability

Series	Month 1	Month 2	Month 3
NAE	4.291	3.800	3.389

Note: The table presents the test statistic from the null hypothesis that the ARDL model and AR model have equal predictive ability. The 5% critical value is 1.645.

Table 7: Forecast Combination Results

Method	Month 1	Month 2	Month 3
ARDL	0.93	0.98	1.01
Cointegration	0.92	1.00	0.97

Note: The table presents the ratio of the RMSEs from disaggregate forecasts relative to forecasting the NAE series directly using the same method. A value less than one indicates that the combination of individual forecasts is more accurate.

Table 8: Direct Monthly Forecasts with Refined Oil Prices

NAE	**Month 1**	**Month 2**	**Month 3**
AO	4.44	6.52	7.57
AR	4.33	6.29	7.12
ARDL	2.90	4.71	6.23
COINT	3.19	5.26	6.91
PETROL			
AO	5.20	7.75	9.09
AR	4.96	7.22	8.19
ARDL	3.51	6.46	8.72
COINT	3.51	6.54	8.49
DIESEL			
AO	3.83	5.77	7.02
AR	3.80	5.75	6.94
ARDL	2.64	4.51	6.09
COINT	2.85	4.94	6.80
FUEL OIL			
AO	7.55	10.13	11.79
AR	7.43	9.76	11.27
ARDL	5.13	7.44	10.96
COINT	4.53	7.97	10.40

Note: The table presents the RMSE for the benchmarks and the specified forecasting approaches.

Chapter 4

Testing for Asymmetric Pricing Behaviour in Irish and UK Petrol and Diesel Markets

This chapter empirically tests whether Irish and UK petrol and diesel markets are characterised by asymmetric pricing behaviour. The econometric assessment uses threshold autoregressive models and a dataset of monthly refined oil and retail prices covering the period 1997 to mid-2009. A methodological note is included on the importance of the specification of the number of possible regimes. In particular, the possibility of conflicting price pressures arising from short-run dynamics in retail prices and responses to disequilibrium errors needs to be explicitly modelled. For both the Irish and UK liquid fuel markets at national levels, the paper concludes that there is no evidence to support the "rockets and feathers" hypothesis that retail prices rise faster than they fall in response to changes in oil prices. It is still possible that a lack of competition at a more local level may accommodate asymmetric pricing behaviour.

4.1 Introduction

The heightened volatility in international crude oil prices in recent years has intensified interest in aspects of the pass-through of oil prices to retail prices in the domestic fuel market. In particular, interest has focused on the speed of the pass-through of oil prices and any possible associated asymmetric pricing behaviour that may suggest retailers have some short-run market power. Generally, when a market is perfectly competitive, the price setting behaviour of firms is symmetric in reaction to increases or decreases of the same order of magnitude in input costs. Although there is evidence to suggest that the profit margins in Irish transport fuel wholesale and retail markets are relatively modest, the perception persists that the retail transport fuel market in Ireland is not entirely competitive with consumers not benefitting from falls in crude oil prices with the same rapidity as they are burdened with rises in crude oil prices. Indeed, the Department of Enterprise, Trade and Employment in Ireland commissioned a study in September 2008 to identify "why pump prices for petrol and diesel have not fallen in line with the drop in the wholesale price of oil".

While the issue of pass-through of oil prices to liquid fuel prices has been treated extensively for the US and larger euro area countries, there is a dearth of studies assessing the oil price pass-through and possible asymmetries in the Irish liquid fuels (petrol, diesel and heating fuel) markets. The National Consumer Agency (NCA) studied the relationship between crude prices and downstream product prices but the main analysis relied on graphical illustrations based on data for 2008 only. The pass-through in Ireland has also been examined by Meyler (2010) using standard asymmetric error correction models as part of study across 12 euro area countries. We adopt a more sophisticated econometric model, the threshold autoregressive model, which we argue has more intuitive appeal. Such a model allows for the possibility of response rates changing when passing a non-zero threshold rather than the typical zero threshold. At the same time, we provide extensive Irish context by, for example, discussing structural indicators for the Irish market. The econometric model is estimated based on a long time series of refined oil and retail (petrol and diesel) prices. Therefore, this paper addresses a significant gap in the literature. An assessment of the structure and pricing behaviour in these markets can also aid short-term inflation forecasting exercises and provide a better understanding of the direct effects of oil price shocks.

Various structural indicators of liquid fuel markets are examined to shed some preliminary light on the issue of a possible lack of effective competition. The main focus of this paper is to determine the pass-through rates and test for asymmetries in Irish fuel markets but the equivalent pass-through rates in the UK are also examined in order to provide some context. It is important to note that, if there are asymmetries, this analysis does not identify why these pricing asymmetries occur or at what stage of the supply chain (wholesale or retail) they may arise. We propose to use a threshold autoregressive model to ascertain whether significant asymmetries characterise Irish and UK liquid fuel markets. Such a non-linear modelling approach has some intuitive appeal, with the higher pass-through kicking in when a certain threshold is passed. While the analysis uses long time series and sophisticated econometric techniques the results on oil price pass-through rates come with some caveats attached. There are challenges to assessing the pass-through of oil prices in the case of Ireland which may limit the robustness of the results. These include the low frequency of the data, which are only available at a monthly frequency whereas a weekly or higher frequency would be preferable. Cross-subsidisation may also distort the pass-through with retailers supporting revenues by increasing margins in non-fuel items to offset tighter margins in pump prices.

The paper takes the following structure. Section 2 briefly discusses the results from previous studies on the oil pass-through to retail pump prices for the Irish and other euro area countries' markets. The subsequent section describes the structure of the fuel markets in Ireland using a range of indicators of competition while also providing some further background information on the pricing mechanisms involved. The data used for the study of pass-through and asymmetries is described in Section 4. Sections 5 introduces the various modelling approaches while Section 6 describes the results. Section 7 presents a methodological note on the importance of a multi-regime specification in threshold autoregressive models. Finally, Section 8 summarises the findings and concludes that broadly there is no evidence to support the view that retail prices rise faster in Irish and UK liquid fuel markets at national levels than they fall in response to oil price changes.

4.2 Oil Price Literature

There are numerous studies of oil pricing asymmetries, forming part of the extensive "rockets and feathers" literature that examines whether retail prices rise faster than they fall in response to changes in oil prices. Many of these studies have been undertaken for fuel markets in the US and larger euro area countries and have adopted a wide range of econometric approaches. Borenstein et al. (1997), in a seminal paper, examined price asymmetries in US gasoline markets and found retail prices showed asymmetry to crude oil price changes, possibly reflecting inventory adjustment effects. Geweke (2004) contains an excellent exploration of the econometric issues faced in analyses of pricing asymmetries in fuel markets and provides a useful critique of the important empirical studies on US fuel markets available up to that point.

Bacon (1991) is one of the first papers to examine the issue of asymmetric gasoline pricing for the UK. The price tranmission process under study is that between international refined prices and retail prices, using bi-weekly data. They find that price increases are transmitted with 2 months but that price decreases appear to require one additional week. The evidence of asymmetery, although present, is quite weak. In a second UK study undertaken the same year, Manning (1991) examines the pass-through from crude oil price to retail prices using monthly data. Again, evidence of asymmetry is weak and non-persistent. With a broader geographical coverage, Manera and Frey (2007) discusses the results from the various empirical studies undertaken in analysing asymmetries in price transmission generally while also providing a comprehensive overview of the alternative econometric approaches adopted. They include the UK in the study and their general finding is that asymmetry of some description can be found for most countries.

There are a limited number of studies that examine the pricing behaviour in Irish fuel markets. The NCA Report investigated the movements of refined prices, wholesale prices and retail pump prices in Ireland during 2008. While the NCA Report concluded that there is little evidence to suggest unwarranted delays in the passing through of wholesale price changes consumer prices, their study was based on a just one year of data and did not assess the issue using econometric models. The study presented valuable information on the working of the liquid fuels industry in Ireland. Meyler (2010) examines the oil pass-through across all euro area countries using a standard asymmetry model and uses levels data rather than the standard approach in the literature of modelling the series in logs. The broad conclusion arrived at in the study is that the pass-through in euro area countries generally is full and quick and there are no significant pricing asymmetries.

The standard asymmetry model may be mis-specified as there is likely to be a fixed cost associated with price changes such that firms make adjustments to prices only when the input cost change is sufficiently large to justify incurring the cost of implementing the price change. A threshold autoregressive model (TAR) may be used to allow for such a non-zero threshold effect i.e. asymmetric pricing is only triggered by a minimum absolute change in crude (or refined) oil costs. The TAR model is a more general form of the standard asymmetry model with the threshold arbitrarily set to zero in the case of the latter. The threshold variable is typically based on the current price change or on the average price change over a number of time periods. These models tend to focus more on the short-run asymmetries - lag to initial response and the duration of the response - rather than the long-run response - whether the cost change is fully passed through. The Hansen (1998) bootstrap approach is commonly used to test the null hypothesis of a linear model against the TAR alternative.

While the TAR approach has much intuitive appeal, the added value from adopting a threshold pricing approach depends on the accuracy of the estimate of the true threshold and on whether the threshold is economically and statistically significantly different from zero. Al-Gudhea et al. (2007) found evidence in support of threshold pricing for the US gasoline market while Asplund et al. (2000) also identified some threshold pricing in Swedish gasoline markets although the fixed costs appeared to be relatively small. Finally, Godby et al. (2000) applied TAR models to the Canadian gasoline markets but did not find any evidence to support pricing asymmetries. The threshold autoregressive model is explained in greater detail in Section 5.

4.3 Structure of the Irish Oil Market

The recent report by the National Consumer Agency provides a detailed examination of the various stages of the oil market. This oil market is currently entirely downstream in nature i.e. activities in the industry are restricted to refining, storage, distribution and sales of refined products. Crude oil is imported to Whitegate, Ireland's only oil refinery, while refined oil is imported to wholesalers' storage depots located around the country. The wholesalers in turn sell to distributors or directly to retailers while the distributors sell to a range of industrial and service station retailers. There are no oil pipelines in Ireland so the product is transported entirely by road or by sea, which can add significantly to costs relative to the UK where there are oil pipelines.

The price of interest in this paper is the price charged to consumers and there is some tentative evidence to suggest that the retail market is relatively competitive. Over 75 per cent of service stations are independently owned but many operate under Solas agreements whereby the retailer is tied to a certain supplier for typically a period of five years. However, these retailers along with service stations licensed by oil companies can set their own prices. In terms of concentration, there are no restrictions on the number of stations. There are up to 2,034 service stations in Ireland and the number of stations per capita in Ireland is one station for every 2,020 inhabitants. This compares with one station per 9,539 inhabitants in the UK and one station per 3,113 people in Northern Ireland. In 2005, the market share of three largest firms in Irish retail transport fuel sector was about 53 per cent while the corresponding share for the euro area was 49 per cent.[1] Turning to the profits of Irish subsidiaries of oil firms, an analysis of profits by the NCA, which looked at the financial accounts of large oil firms, suggests that net profits after tax as a proportion of the cost of sales (including from non-fuel sales) are quite modest at between 0.6 per cent and 2.3 per cent. Altogether, this suggests that the market structure at national level is comparatively competitive although fuel markets are essentially local in nature and the degree of competition may vary significantly across regions, cities and towns.

While the frequency of price changes in the wholesale market can be quite high, the prices often tend to be based on monthly averages of refined oil prices. About 30 per cent of the oil imported into Ireland is destined for the Whitegate oil refinery, with the majority of customers (about 60 per cent) of Whitegate paying monthly averages for refined oil while the remaining customers pay based on twice weekly rates. Wholesalers that purchase imported refined oil mainly pay at a price based on the average of the month and tend to change their selling prices twice a week based on the average of refined prices over the previous two or three days. Currency and other hedging may also dampen fluctuations in the wholesale price. As a result, any slow speed of pass-through of refined prices to wholesale prices may partly reflect a degree of smoothness in prices due to trading on futures markets although it would not explain any persistent asymmetries in pricing. The frequency of price changes at the retail

[1] The corresponding figure for the UK is not available.

stage is also quite high. The pricing mechanism for fuel products may be distorted somewhat as retailers may avail of opportunities for cross-selling with higher margins in non-fuel items, such as food and beverages, compensating for tighter margins in petrol and diesel. Moreover, the entry of large supermarket chain such as Tesco into the petrol and diesel retail markets may have accelerated this trend, although the impact is likely limited to a relatively small number of localities for now, particularly given that retail size restrictions remain in place.

4.4 Data

There are a number of choices to make in terms of the data for a study of this nature. For the international oil price, one can use the price of either crude or refined oil. The prices of refined gas and diesel are used in this study, as refined prices reflect the cost to the wholesaler or retailer more closely than crude prices. The refining margins have fluctuated significantly in recent years, particularly for diesel, and therefore using crude prices are not an appropriate proxy for the prices ultimately paid by the wholesaler or retailer. The refined oil prices are Rotterdam gasoline and diesel prices in dollars per barrel converted to euro per litre. Regarding the retail price, again there are two options - including the pre-tax or final tax-inclusive retail price in the study. The choice of retail price depends on the research question. The emphasis in this paper will be on the pass-through to the pre-tax price, as this price is controlled by retailers. The monthly pre-tax petrol and diesel price level data are taken from the European Commission Weekly Oil Bulletin and unpublished monthly price indices for petrol and diesel are kindly provided by the CSO.

All data are at a monthly frequency and at national level. Ideally, the data would be at a weekly or higher frequency and available at local level rather than national. There are higher frequency data available on user websites such www.pumps.ie but they are of questionable reliability. There are also wholesale price data available from www.fuelsonlineprices.com. Any finding of asymmetry between refined oil price changes and retail price changes could be further investigated in a follow-up study on the wholesale and retail stages separately to identify at which stage it may be more of an issue. In this respect, it is interesting to note that the margins seem to be of a similar order at both stages - around 2.5 cents per litre. The monthly prices data for the UK have also been taken from the EC Weekly Oil Bulletin. The data collection methodologies for Irish and UK data may differ somewhat so comparisons may not be strictly comparable. Prices in the fuel supply chain in Ireland are often based on monthly averages, which suggests that the pass-through in the case of Ireland may be relatively sluggish and any significant asymmetries may still be picked up using data of a monthly frequency.

4.5 Estimating Asymmetry

4.5.1 Basic Cointegration Model

The price of liquid fuels at retail level closely follows developments in refined prices. Figure 1 shows the price of refined gasoline together with the retail price of petrol. The strong observed comovement between retail and refined prices is suggestive of a cointegrating relationship between the two. Figure 2 shows the logs of these prices, with separate scales this time. Again, there is strong comovement with the log series.[2] This implies a long-run relationship of the form

$$PC_t = \alpha + \phi PR_t + \epsilon_t \tag{4.5.1}$$

where PC_t is the consumer or retail price at time t and PR_t is the refined price. Augmented Dickey-Fuller tests find that all the price series contain a unit root regardless of whether we use logs or levels. Using the Engle-Granger two-step approach, the residual of the long-run equation is checked for a unit root and the null of no cointegration between refined and retail prices is very strongly rejected in each case[3] and so our initial statistical tests confirm the commonly found cointegrating relationship between retail and refined prices. This relationship allows us to embed the long-run residual, which is the ECM term $\epsilon_t = ecm_t = PC_t - \phi PR_t - \alpha$ in an error correction model (ECM) of the form:

$$\Delta PC_t = \gamma + \theta ecm_{t-1} + \sum_{i=1}^{q} \eta_i \Delta PC_{t-i} + \sum_{j=0}^{p} \beta_i \Delta PR_{t-i} \tag{4.5.2}$$

The existence of a cointegrating relationship between refined and retail prices for both logs and levels means we have a choice in terms of how to specify the model. Although specification in terms of logs is standard in most econometric models, we choose the levels specification for both statistical and conceptual reasons. Statistically, the long-run equation shows much stronger cointegration properties when specified in levels rather than logs. Furthermore, in a study of European fuel prices, Meyler (2010) finds that there can be instability in the long-run relationship if specified in logs but this problem does not arise in the levels specification. The final statistical benefit of the levels specification is in terms of the fit of the short-run equations. We find there is considerable improvement in the R^2 of the short-run equation when the specification is in levels.

It also makes more sense theoretically to specify the model in terms of absolute price levels because a one cent increase in refined prices will lead to a one cent increase in pre-tax prices if there is full pass-through. However, the impact of a one per cent increase in refined prices on post-tax prices will depend on the actual level of refined

[2]The graphs for diesel show similar properties.

[3]The results of unit root tests and cointegration tests are not presented in the interests of brevity but are available from the authors upon request.

and retail prices. There is a large tax element to retail prices. This mainly consists of excise duty, which is a fixed tax, and VAT, which is an ad valorem tax. The tax element accounts for a larger percentage of the retail price when prices are low and a smaller percentage when prices are higher. Therefore, if we model in logs, pass-through will depend on the price level. This can be avoided by using a model in levels. For this reason, together with the superior statistical properties, the log specification is dropped in favour of the levels specification for the rest of the analysis. The levels specification could result in a problem with heteroscedasticity so all results presented later are based on Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors.

From the equations above, it is implied that the cointegration relationship is estimated using the Engle-Granger two-step procedure. In the interests of robustness, we also use the ARDL approach suggested by Peseran, Shin and Smith (2001). This approach uses a one-step non-linear estimator, in which the long-run cointegrating relationship is embeded in the short-run equation. In Table 1, we present the estimates of the intercept, α and the slope coefficient, ϕ, from the long-run equation for both the ARDL non-linear estimator and the OLS estimator of the Engle-Granger two-step procedure. The ARDL method and the OLS method yield very similar results for both petrol and diesel in both the UK and Irish markets. Although not formally tested, it is clear that the differences in the parameter estimates are well within the bounds of statistical variation. This fact there is very little difference between the two estimators suggests that the results of the study are not sensitive to the choice of estimator. For this reason, we choose to conduct the remainder of our study using the two-step estimator.

4.5.2 Simple Models of Asymmetry

The standard cointegration model is not sufficiently rich to capture the underlying dynamics if there is asymmetry in the transmission of changes in the price of refined oil[4] to retail prices. The model presented in equation 4.5.2 implicitly assumes symmetric adjustment around $\Delta PR_t = 0$, so that all coefficients are the same regardless of whether changes in refined prices are positive or negative. To allow a different response to positive and negative refined price changes, define the following indicator variable, I_t:

$$I_t = \begin{cases} 1 & \text{if } TR_t > \tau \\ 0 & \text{if } TR_t \leq \tau \end{cases}$$

where TR_t is the value of the threshold variable at time t and τ is the estimate of the threshold. In the zero threshold model, the threshold variable is the change in the

[4]We use the general term "refined oil" instead of referring to both refined gasoline and refined diesel.

refined price, ΔPR_t, and τ is zero. The model can then be expressed as:

$$\Delta PC_t = \gamma + \theta_1 ecm_{t-1} + \sum_{i=1}^{q} \eta_i \Delta PC_{t-i} + \sum_{j=0}^{p} \beta_{1,i} \Delta PR_{t-i}, \quad I_t = 1 \qquad (4.5.3)$$

$$\Delta PC_t = \gamma + \theta_2 ecm_{t-1} + \sum_{i=1}^{q} \eta_i \Delta PC_{t-i} + \sum_{j=0}^{p} \beta_{2,i} \Delta PR_{t-i}, \quad I_t = 0 \qquad (4.5.4)$$

One set of parameters apply when the change in the refined price is positive and another set when that change is negative. Although the model is presented as two separate equations to clarify the concept, estimation proceeds in a single equation manner. The equations above can give rise to a number of different types of asymmetry and we consider three alternatives. We first estimate a model in which the only source of asymmetry is through different parameters for the ECM term. The speed of adjustment towards equilibrium varies depending on the value of the threshold variable.[5] In terms of the coefficients, this model means that $\theta_1 \neq \theta_2$ but the other coefficients in the model are restricted to be the same in both regimes. In the second type of model, we only allow asymmetry in the lags coefficients on the refined price. In this scenario, the ECM coefficient is the same in both regimes. Thus the $\beta_{1,i}$ and $\beta_{2,i}$ coefficients vary but the other coefficients are constant. Finally, we allow both the ECM and lag parameters to vary. This is the most flexible model but the lags on the autoregressive terms and the constant are still invariant across regimes. We do not allow the autoregressive parameters to change because the price pressures in the system should come through the equilibrium error and the change in the upstream price. Although it is conventional to follow this approach in the literature, the assumption that the autoregressive error is the same in both the symmetric and asymmetric cast is not formally tested in the paper. This could be problematic if the persistence of the series differs in the different regimes but we find that the autoregressive term often drops out of the specification in the asymmetric model. This suggests that once the correct model is found, changes in the refined price and equilibrium error are the main sources of retail price dynamics.

In the preceeding discussion, the threshold was defined to be the change in the refined price. We also allow for alternative threshold variables. In particular, we allow the lagged change in the refined price and a moving average of recent price changes as two alternative threshold variables. The moving average is based on the current change and the lagged changes from the past two months. For each of the three threshold variables, we estimate the three models of asymmetry described above to give a total of nine models of the zero threshold variety. We present and discuss results in a later section once all asymmetric models have been explained.

Figure 3 presents a graphical representation of a zero threshold model. The graph shows the change in retail prices on the Y axis for a given change in the refined

[5] Although asymmetric responses in the ECM are permitted, the long-run cointegrating relationship between the refined price and the retail price represented in equation (1) must still be identical for price increases and decreases. This restriction is required to prevent the possibility of ever widening margins.

price on the X axis. This is for illustration purposes so it only represents the short-run response. The black line represents the symmetric case. In this hypothetical economy, half the change in the crude price is reflected in the retail price in the symmetric case. The dashed line represents the asymmetric case. When the change in the crude price is negative, one quarter of the change is reflected in the retail price. The diminished responsiveness relative to the symmetric case is evident from the fact that the black line is closer to the X axis when below zero. If the change in the refined price is positive, the change in the retail price is more responsive for the asymmetric model with three quarters of the crude change passed through to refined prices. Although simplistic by design, this example highlights the nature of a model with a zero threshold.

4.5.3 Interpreting the Model

The specification presented through equations 4.5.3 and 4.5.4 is the basic approach to specifying asymmetry in a cointegration framework. Parameter change models based on indicator variables are quite standard tools in econometrics. However, one needs to be very careful in terms of the interpretation of this model. In this section, we present the results for the long run equations for the pre-tax petrol and diesel models. This gives us an insight into the behaviour of the equilibrium error, which then allows us to describe the model more intuitively.

Returning to Table 1, which presents the estimates of the coefficients of the long-run equations, consider the OLS results for the Irish models. The coefficient on the refined price, ϕ, is close to one in both cases. This shows that movements in the refined price are matched almost one for one with movements in the retail price. The intercept, α, in these equations is an estimate of the equilibrium margin between retail and refined prices. This is approximately 17 cent for petrol and 19 cent for diesel in the sample. The equilibrium error has already been expressed as $ecm_t = PC_t - \phi PR_t - \alpha$. Given that $\phi \approx 1$, this equilibrium error is the current margin between retail and refined prices less the equilibrium margin.

The intuition becomes more obvious when we examine Figure 5, which depicts both the margin between retail and refined prices and the equilibrium error. The graph shows that when the equilibrium error is positive, the margin between retail and refined prices is above its long-run average with the corollary true when the equilibrium error is negative. Given the coefficient of close to one in the long-run equation, changes in the equilibrium error largely reflect changes in the retailer's margin. Retail prices lag refined prices under the assumed pass-through relationship. It follows that the equilibrium error will be positive, and so the margin above its long-run average, when there has previously been a fall in the refined price. This is borne out by the data, which show a large negative correlation between the value of the equilibrium residuals and the change in the refined price. Taking the Irish data for example, the correlation is -0.82 in the case of petrol and -0.78 for diesel.

If retailers allow prices to rise more quickly than they fall, as posited in the rockets and feathers literature, this will be represented by a larger coefficient on the positive ECM term. Positive equilibrium errors typically correspond to a period of falling refined prices. The speed of adjustment coefficient is always negative as it is working to bring the system back to equilibrium. In an asymmetric model, a smaller coefficient on the positive ECM term relative to the negative means that there is less downward pressure on retail prices to restore the equilibrium margin after a period of falling refined prices compared to the upward pressure on retail prices following a period of increasing refined prices. This is the type of asymmetry we search for in the ECM coefficients.

We have mentioned three possible threshold variables for the paper - the current, lagged and moving average change in the refined price. The approach described in equations (3) and (4) allows certain coefficients in the model to vary depending on the value of the change in the refined price and this approach is often adopted in the literature. The value of the equilibrium error is another source of price pressure in the system. A model which divides the sample according to positive or negative values of the threshold variable does not result in a corresponding division of the equilibrium error into positive and negative values. There may be conflicting price pressures from the change in the refined price and the value of the equilibrium error. It is readily apparent that four possible cases arise.

Interactions between ECM and short-run dynamics

	ECM(+)	ECM(-)
$\Delta PR(+)$	(1) + +	(2) + -
$\Delta PR(-)$	(3) - +	(4) - -

In the first regime, both the equilibrium error and the change in the upstream price are positive. A positive equilibrium error means that retail prices are above their equilibrium level so there will be downward pressure on the retail price. The change in the retail price and the error correction mechanism will work against each other in this situation. In regime 2, both the positive change in the refined price and the negative equilibrium error exert upward pressure on retail prices and in regime 3 they both exert downward pressure. Regime 4 again represents a scenario where there are opposing price pressures.

We believe that the comparison between regimes 2 and 3 is the most valid, as there are no conflicting price pressures. The model as presented in equations (3) and (4) partitions the sample according to positive and negative changes in the refined price. However, when the change in the refined price is positive, the equilibrium error can still take on positive values, putting opposing pressure on price. If we wish to distinguish between the behaviour of retailers when price pressures are unequivocally positive or negative, we must compare regime 2 with regime 3. For this reason, we estimate threshold models with four different regimes and the ECM and lag coefficients can vary in each regime.

4.5.4 Threshold Models of Asymmetry

It may not be the case the threshold is actually zero in practice. Retailers may change their behaviour when the change in the refined price, be it positive or negative, passes some threshold value other than zero. A non-zero threshold could arise for cost and competitiveness reasons or simply as a result of strategic pricing. In this section, we consider non-zero threshold models of asymmetry. It is possible to test for non-zero thresholds using a procedure developed by Hansen (1998).

The threshold model extends the model presented above in a natural way. Instead of defining the indicator variable according to values of the threshold variable above and below zero, a non-zero value of τ is now permitted as the threshold estimate. Figure 4 provides a graphical representation. In the example shown, the behaviour of the dashed line, which is indicated by its slope, is the same for negative and small positive changes in the threshold. However, for larger positive changes in the refined price, the responsiveness of the retail price increases dramatically, as represented by the large jump in the graph.

In terms of implementing this approach, the candidate threshold variable TR_t is ordered by size and 15% of the tails are trimmed. For each possible remaining value of the threshold, a model like that in equation (3) is estimated. The residual sum of squares and threshold value is recorded for each model. The model with the lowest residual sum of squares is taken as the estimate of the threshold. Hansen (1998) provides a bootstrap procedure to test for the statistical significance of the threshold estimate and we use 5000 replications in our test.[6] As with the zero threshold models, we consider three threshold variables and three types of asymmetric model meaning that a total of nine models are estimated. Standard model selection criteria are used to choose between them. For each model, we calculate the Akaike Information Criteria (AIC) and the Schwarz Bayesian Criteria (SBC) and the appropriate model is chosen based on consideration of these two statistics in conjunction with the adjusted R^2.

It is equally appropriate to consider the four different regimes when implementing the non-zero threshold. The threshold procedure trims 15% of the tails. If the threshold value chosen by the procedure is close to the trimmed portion, there is a relatively small number of observations above the threshold value. As the values lying above the threshold also need to be divided according to positive and negative values of the equilibrium error, there may not be enough observations in one or both regimes to get accurate parameter estimates but it is possible to deal with problem. Generally, most

[6] For further details of the testing procedure, see Hansen (1998).

observations that are above threshold are associated with a negative equilibrium error, which represents the unequivocally positive case (Regime 2). As there are so few observations for the regime with positive equilibrium errors and above threshold, this regime is merged with the regime corresponding to negative equilibrium errors and below threshold. In other words, the two regimes with conflicting price pressures are collapsed to one regime with comparison still valid between the wholly positive and negative regimes. One would expect the regimes with conflicting price pressures to have more sluggish adjustment than regimes with reinforcing price pressures a prori. This should also hold true for the merged conflicting regime. In any case, the merger of these regimes should not impact the comparison between the positive and negative regimes, which is the topic of interest.

4.6 Results

4.6.1 Asymmetry in the Irish Market

In Table 2, we present the model selection criteria for both pre-tax petrol and diesel models. The AIC and SBC are calculated such that a smaller value indicates a better fit. Models are chosen based on both the R^2 and the model selection statistics. The statistics do not always choose the same model and in this type of scenario the is chosen based on the balance of evidence from all three statistics. Results are presented for a standard ECM model with no asymmetry, an asymmetric ECM model, a zero threshold model and a non-zero threshold model in the cases where a non-zero threshold exists. These terms are explained in the previous section but a glossary of model descriptions is provided at the back of the paper for convenience. The table shows the model, the type of asymmetry allowed, the threshold variable and the model selection statistics. The specification column is mainly for use with the threshold models and indicates if asymmetry is permitted in the ECM term, the lags or both the ECM and the lags. For each type of model, results are presented for the model with the best diagnostics out of the nine possible combinations of threshold variable and asymmetry type.

The first row in each section of Table 2 presents the results for the basic model with no asymmetry. Subsequent rows provide the same statistics for asymmetric models of the description provided. In both cases, a three regime model which allows asymmetric behaviour in the lag response is selected. Table 3 presents the coefficients for the chosen petrol and diesel model. The "2" superscript on the lag coefficients denotes the coefficients for the regime with positive price pressures, the "3" superscript applies to the regime with negative price pressure and the lags without a superscript refer to the regime with conflicting price pressure. For petrol, the ECM coefficient indicates that 43% of any equilibrium error is eliminated in the following period, implying quite a fast speed of adjustment. If we sum the coefficients on the lag terms for each regime, we get a total of 0.614, 0.473 and 0.662 for the positive, conflicting and negative regimes respectively. The coefficient sums are larger for the outer regimes so that adjustment is quicker when the price pressures are persistently in either the positive or negative direction. However, the summed coefficients are very similar for the positive and negative regimes. We test the null hypothesis that these coefficient sums are not equal. As the residuals are non-normal, we use a bootstrap procedure with 5000 replications to construct the test statistic and reject the null hypothesis that the coefficient sums are not equal. There is no evidence that pass-through is greater for positive changes in the refined price. In the case of negative price pressures, the effect is contemporaneous in the sense that adjustment takes place in period t. For the positive regime, the bulk of the adjustment takes place in $t-1$. Thus, although the magnitude of adjustment is the same for the positive and negative regimes, adjustment is marginally quicker for the negative regime.

A similar pattern emerges in the diesel market. The ECM coefficient is almost -0.7, indicating that a lot of the adjustment in the model is taking place through the error correction mechanism and the coefficients on refined prices changes are correspondingly smaller. The summed coefficients are 0.404, 0.159 and 0.532 for the positive, conflicting and negative regimes respectively. Once again, the conflicting regime has a more sluggish response. The reaction to negative price changes is again contemporaneous. For positive price changes, half of the pass-through takes place in period $t-2$. The negative pass-though coefficient is a bit stronger than the sum of the positive but the null hypothesis that they are not equal is rejected at the 5% level. Again, there is no evidence in favour of the perception that prices increase faster than they fall. The results of these models indicate that there is asymmetry in the market, as asymmetric models outperform symmetric models. The retail price response is faster when price pressures are either wholly positive or negative. In addition, the negative response appears to be marginally quicker than the positive response. The retail of petrol and diesel is a very competitive industry in Ireland and is characterised by tight margins. The asymmetry found in the models is consistent with this type of market.

4.6.2 Asymmetry in the UK Market

Table 4 provides an overview of the model selection criteria for the UK. The results for the UK refer to pre-tax prices again. For petrol, the statistical evidence is in favour of the non-zero threshold model as the R^2 and AIC statistics are slightly better although the SBC chooses the standard symmetric model. Thus, we opt to estimate the asymmetric model and check for asymmetry but we stress that the improvement in fit over the symmetric model is very small. For diesel, the evidence is more clear-cut as all selection statistics point to the zero threshold model.

Table 5 presents the coefficients for the selected models. In the case of petrol, the asymmetry in the model comes from different coefficients on the ECM terms in the three different regimes. The coefficients on the ECM terms are -0.493, -0.434 and

-0.233 for regime 3, the neutral regime and regime 2 respectively. Regime 2, the positive regime, has a change in the refined price which is above threshold and a negative equilibrium error. For this negative equilibrium error, the speed of adjustment coefficient of -0.233 means that upward pressure on prices eliminates nearly 23% of any equilibrium error in the following period. Regime 3, which is the negative regime, has a change in the refined price which is below threshold and a positive equilibrium error. In this regime, the speed of adjustment coefficient means that the downward pressure on prices eliminates approximately 49% of any equilibrium error in the following period. In the case of conflicting price pressures, the speed of adjustment parameter is about 43%. We saw that the evidence in favour of the asymmetric model is very weak for petrol and the results show that, if it does exist, asymmetry is working in the opposite direction to that suggested by the "rockets and feathers" literature.

For the diesel model, asymmetry is permitted in both the ECM term and the lags. We see that the coefficient on the ECM in regime 3 is again greater than in regime 2, again suggesting weaker error correction in the upward direction. However, the difference between the two is not statistically significant according to the bootstrap test. The ECM term for the regime with conflicting price pressure is not statistically significant. The summed coefficients on the refined price terms are 0.763, 0.437 and 0.917 for the positive, conflicting and negative regimes respectively. As with the Irish results, the parameter sums are greater when price pressures are unambiguously positive or negative. Although the response is slightly stronger for the negative regime, the positive and negative parameter sums are not statistically different. Thus, the only statistically significant asymmetry is between conflicting and reinforcing regimes. When we compare positive and negative price pressure, the coefficients are somewhat stronger in the negative direction, as we found with diesel, but the differences are not statistically significant. In any case, we can reject the "rockets and feathers" hypothesis. The rejection of this hypothesis is at odds with some of the empirical literature but, as mentioned in the literature review, the evidence of asymmetry in the UK has often been quite weak.

The asymmetry in the UK markets for petrol and diesel differs to the extent that the asymmetry comes through the error correction mechanism for petrol but through the direct response to changes in refined prices for diesel. When we consider the UK and Irish results in tandem, the asymmetry and its specification is similar. For both Irish markets and the UK diesel market, the asymmetry is manifested through the coefficients on the lagged changes in refined prices. The improvement in fit for these models relative to the symmetric model may be modest but it is clear that asymmetry exists. The price response is stronger for the regimes with reinforcing price pressures. Furthermore, when we compare the positive and negative regimes, there is evidence of modest asymmetry in favour of the negative direction. These results show no support for the contention that retailers respond more quickly to increases in the international price of oil.

As a robustness check, we also estimate the zero threshold asymetric models on weekly UK data.[7] In the case of petrol, a four regime model with asymmetric lags responses is selected. There is a stronger lag response when price pressures are either entirely positive or entirely negative relative to the regimes with conflicting price pressures. However, these positive and negative lag responses are of equal magnitude. For the diesel model, the symmetric model has the best fit according to the AIC and SBC. There are some asymmetric models for which the adjusted R^2 is marginally higher and for these models there are stronger coefficients on negative lags. On the balance of evidence however, the symmetric model is chosen over these asymmetric models. Thus, for the weekly data, there is no evidence in support of the "rockets and feathers" literature.

4.7 Importance of Multi-Regime Specification

While some papers in the literature acknowledge that there may be conflicting price pressures from refined price changes and the equilibrium error, the models are still restricted to two regimes based on the behaviour of some threshold variable or occasionally in terms of the equilibrium error. There are some instances of three regime models but the three regimes are defined according to the behaviour of one threshold variable. We believe that that it is important to consider the price pressures from both the ECM term and a threshold variable through a multi-regime model of the sort used in this paper.

As a counterfactual exercise, we take the models chosen earlier with the best model selection statistics and compare these models to the best performing two regime models. The two regime models are chosen according to the same methods used for the three regime model. The regimes are defined according to whether the refined price threshold variable is above or below its threshold value and so no account is taken of whether the equilibrium error is positive or negative. Table 6 presents the model selection statistics for these two regime models. For convenience, it also contains the statistics for the symmetric models and the corresponding three regime models reported in earlier tables.

With the exception of the UK petrol market, the two regime model is generally outperformed by the three regime model. Thus, in most cases in this paper, the division of samples into periods of wholly positive or negative price pressures helps to improve the fit of the model. The two regime models outperform the symmetric models on the balance of evidence but the improvement in fit is generally small. Ostensibly, it appears that both the two and three regime models find the same asymmetry but that the more refined three regime model picks it up more clearly. Closer examination of the two regime models shows that this is not always the case. For Irish petrol prices,

[7] Tables of results are not provided in the interests of brevity. Non-zero threshold models are not considered.

the two regime model is virtually identical to the three regime model.[8] For diesel, the asymmetry in the two regime model comes from a larger coefficient for one of the ECM terms. However, the bootstrap procedure shows the ECM coefficients are not significantly different so the model collapses to the symmetric case. Thus, for the Irish diesel market, we would conclude there is no asymmetry if using two regime threshold models.

In the case of the UK petrol market, asymmetry in the three regime model is found in the ECM terms, with upward pressure weaker than downward pressure. The selected two regime model also finds some evidence of asymmetry in the ECM coefficients and the bootstrap procedure confirms that the difference in the ECM coefficients is statistically different at the 5% level. In contrast to the three regime case, it is the negative ECM term which is greater for the two regime model and so the results are completely different. As the ECM terms are now defined according to the threshold variable, this means that there is stronger error correction when the change in the refined price is negative. This case highlights most clearly the problem with allowing different ECM responses based only on another threshold variable - it is very difficult to interpret the meaning of the ECM coefficients because the equilibrium error could be positive or negative when the refined price change is below threshold.

For the UK diesel market, the best performing two regime model is identical to the three regime model. However, the model selection statistics do not provide strong evidence in favour of the two regime model over the symmetric model. As a result, the symmetric model is likely to be chosen when in fact an asymmetric model is required. This exercise demonstrates that two regime models based on a threshold variable can lead to significantly different results to those from three regime models (or four regimes where the data permit). In two of the four markets, we may have concluded there is no asymmetry using conventional models and in third market, we would have found asymmetry of the opposite direction. As the model selections statistics show that multi-regime models have the best performance, it is critical to divide observations into periods of wholly positive or negative price pressures when assessing asymmetry. Given that the literature is almost entirely dependent on two regime models, it raises serious questions regarding existing findings of asymmetry in other markets.

4.8 Summary and Conclusion

A range of structural indicators suggest that Irish retail fuel markets at national level are comparatively competitive, with the markets characterised by relatively tight margins. Still, the perception persists that the retail transport fuel market in Ireland may not be entirely competitive with consumers not benefitting from falls in crude oil prices with the same rapidity as they are burdened with rises in crude oil prices. This perception is to a large extent based on anecdotal evidence only, as there is a dearth

[8]Tables outlining the ceofficients of the four 2 regime models are not provided in the interests of brevity.

of studies that have empirically tested whether Irish fuel markets are characterised by asymmetric pricing behaviour. The NCA investigated the movements of refined prices, wholesale prices and retail pump prices in Ireland during 2008. While the NCA concluded that there is little evidence to suggest unwarranted delays in the passing through of refined oil price falls to consumer prices, their study was based on a just one year of data and did not assess the issue using econometric models. In contrast, this paper contains a rigorous econometric assessment of the pass-through of oil prices in Irish liquid fuel markets using a long span of data. The UK fuels market is also examined for comparison purposes.

The econometric analysis uses threshold autoregressive models to test for asymmetries in the Irish fuel markets. The paper contains a methodological note on the importance of allowing for 3 or 4 regimes in the threshold model specification rather than the two regimes that is standard in the literature. The extension to more than two regimes recognises the conflicting price pressures that may arise from short change dynamics and disequilibrium errors. This paper demonstrates that two regime models for a given threshold variable can lead to significantly different results to those for three regime models (or four regimes where the data permit). In two of the four markets, it may have been concluded there is no asymmetry using conventional models and in third market, asymmetry of the opposite direction would have been found. As the model selection statistics show that multi-regime models (i.e. more than two regimes) have the best performance, it is important to divide observations into periods of wholly positive or negative price pressures when assessing asymmetry. Given that the literature typically relies on two regime models, it raises questions regarding existing findings of asymmetry in other markets.

For the Irish petrol and diesel markets, the pass-through of refined oil price falls appears to be more immediate when price pressures are unambiguously negative. This is in contrast to the popular perception that retailers respond more quickly when prices are positive. The asymmetry in the UK markets for petrol and diesel differs to the extent that the asymmetry comes through the error correction mechanism for petrol and through the direct response to changes in refined prices for diesel. When we consider the UK and Irish results in tandem, the model specifications and the findings of asymmetry are broadly similar. For both Irish petrol and diesel markets and also the UK diesel market, there is a faster response to declines in refined prices when price pressures are unambiguously negative. This is a somewhat unexpected result but similar findings can be been found in the literature on asymmetries for the UK fuel market. Also, while such asymmetry may be statistically significant, it is unlikely to be economically significant. More to the point, there is clearly no evidence in either the Irish or UK liquid fuel markets to support the "rockets and feathers" hypothesis that retail prices rise faster than they fall in response to changes in oil prices at a national level. This does not exclude the possibility that at a more local level a lack of competition may accommodate asymmetric pricing behaviour.

4.9 Figures and Tables

Figure 1: Consumer Price and Refined Pices

Euro Cents per Litre

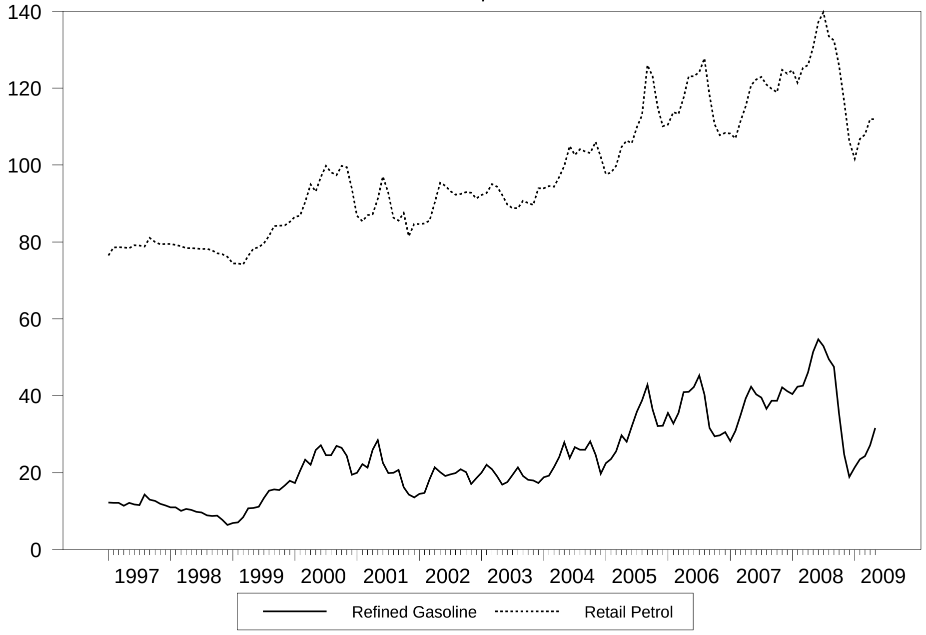


Figure 2: Consumer Price and Refined Pices

Logs

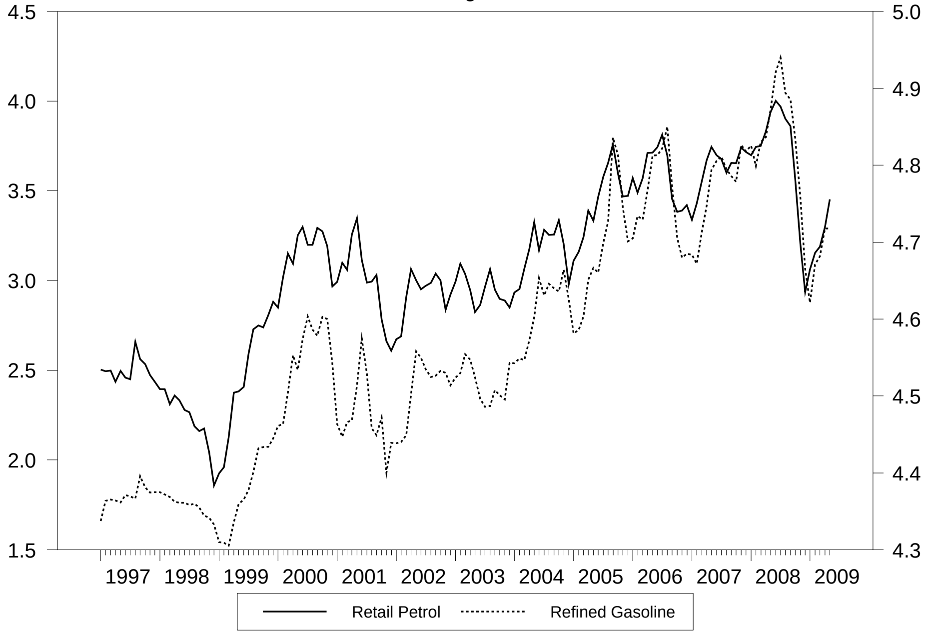


Figure 3: Model with Threshold at Zero

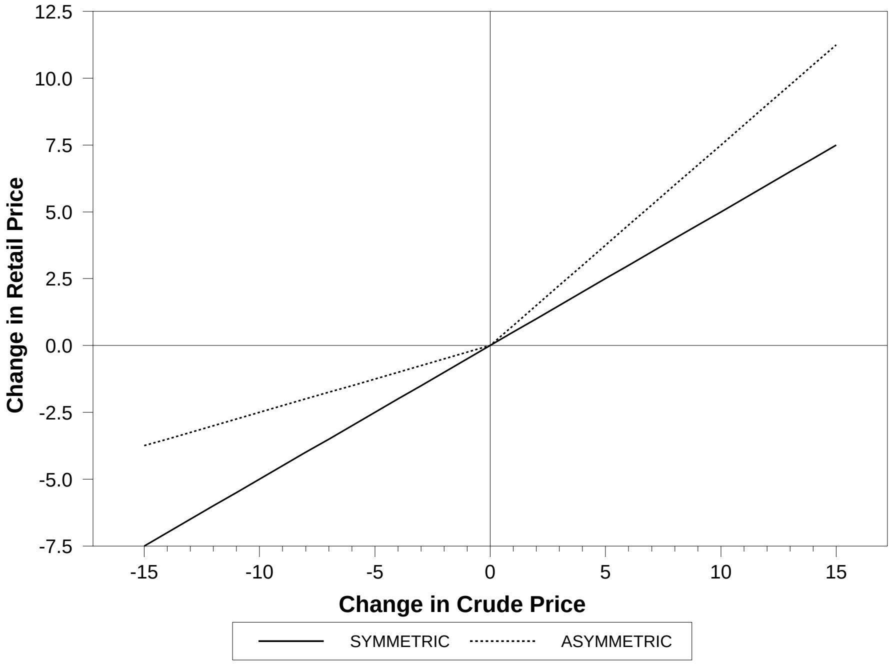


Figure 4: Model with Non-Zero Threshold

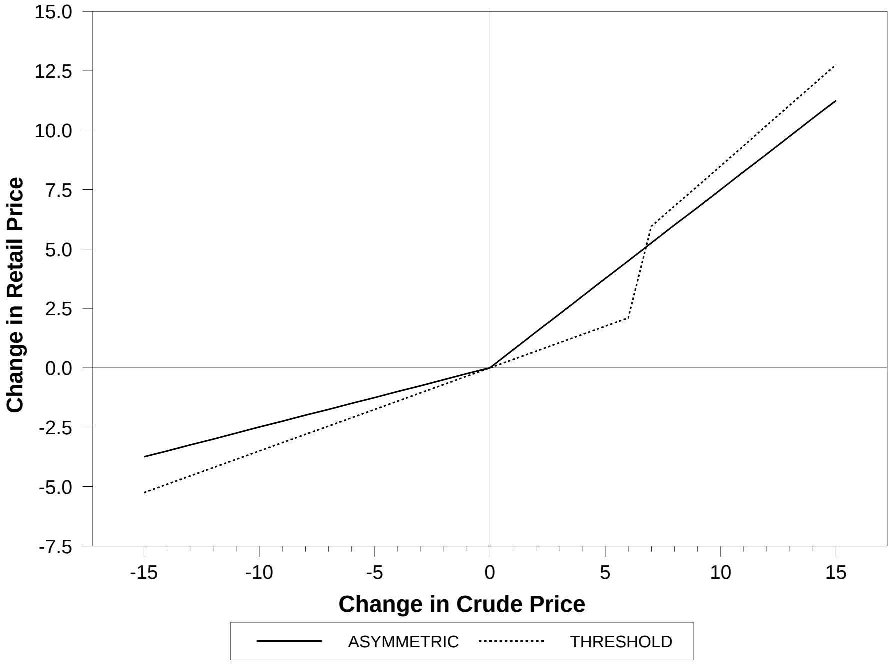


Figure 5: Retail-Refined Spread and Equilibrium Error for Petrol

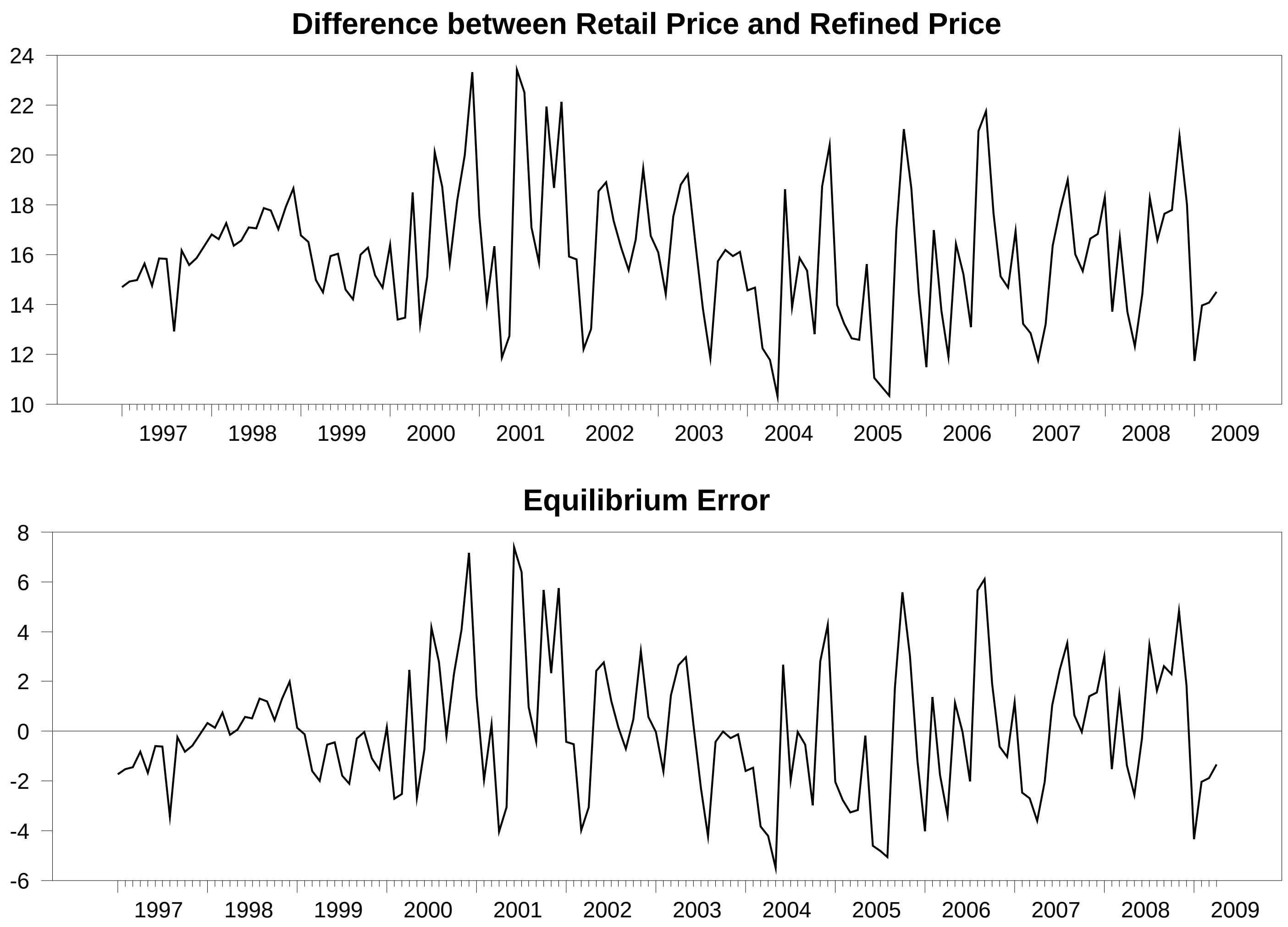


Table 1: Coefficients of Long Run Cointegration Equations

	Petrol Ireland		**Diesel Ireland**	
Regressor	ARDL	OLS	ARDL	OLS
α	16.856	16.920	18.753	19.175
Std. Error	0.610	0.511	0.377	0.441
ϕ	0.967	0.960	1.012	0.993
Std. Error	0.023	0.019	0.013	0.015
	Petrol UK		**Diesel UK**	
Regressor	ARDL	OLS	ARDL	OLS
α	7.864	7.927	9.831	9.705
Std. Error	0.505	0.394	0.982	0.402
ϕ	1.009	1.004	1.019	0.993
Std. Error	0.019	0.015	0.035	0.015

Note: The dependent variable is the retail price and the independent is the refined price. All coefficients are significant at the 5% level. The ARDL columns refers to coefficients estimated using the Peseran et al (2001) one-step non-linear approach while the coefficients in the OLS columns are estimated using the two-stage Engle-Granger OLS method.

Table 2: Overview of Model Performance for Ireland

PETROL

Model	**Type of Asymm**	**Thresh. Var**	**AIC**	**SBC**	R^2
Symmetric	N/A	N/A	3.479	3.603	0.787
Asymmetric ECM	ECM Only	N/A	3.488	3.631	0.787
Zero Threshold	Lags Only	Current	3.364	3.508	0.811
Threshold = -0.4c	ECM and Lags	Moving	3.421	3.606	0.803

DIESEL

Model	**Type of Asymm**	**Thresh. Var**	**AIC**	**SBC**	R^2
Symmetric	N/A	N/A	3.644	3.748	0.714
Asymmetric ECM	ECM Only	N/A	3.632	3.757	0.719
Zero Threshold	Lags Only	Moving	3.532	3.655	0.741

Note: Please refer to the first paragraph of section 4.6.1 for the description of this table. In addition, a glossary of model descriptions is provided at the back of the paper.

Table 3: Coefficients for Selected Model Specification for Ireland

	Petrol		Diesel	
Regressor	Coeff.	Std. Error	Coeff.	Std. Error
Constant	0.146	0.088	0.100	0.109
ecm_{t-1}	-0.434	0.087	-0.699	0.070
ΔPR_t^2	0.212	0.127	0.198	0.115
ΔPR_{t-1}^2	0.402	0.113		
ΔPR_{t-2}^2			0.206	0.107
ΔPR_{t-1}	0.316	0.082		
ΔPR_{t-2}	0.163	0.030	0.159	0.068
ΔPR_t^3	0.662	0.106	0.532	0.082

Note: The dependent variable is the change in the retail price. ΔPR_t is the change in the refined price. The superscript "2" refers to regime 2, the superscript "3" refers to regime 3 and no superscript for the other regime. All coefficients significant at 10% level except for constant in diesel equation.

Table 4: Overview of Model Performance for the UK

PETROL

Model	**Type of Asymm**	**Thresh. Var**	**AIC**	**SBC**	R^2
Symmetric	N/A	N/A	-6.377	-6.275	0.860
Asymmetric ECM	ECM only	N/A	-6.377	-6.255	0.860
Zero Threshold	ECM only	Moving	-6.377	-6.235	0.861
Threshold = 0.5c	ECM Only	Moivng	-6.382	-6.239	0.862

DIESEL

Model	**Type of Asymm**	**Thresh. Var**	**AIC**	**SBC**	R^2
Symmetric	N/A	N/A	-6.014	-5.912	0.768
Asymmetric ECM	ECM only	N/A	-6.008	-5.886	0.768
Zero Threshold	ECM and Lags	Current	-6.076	-5.914	0.786
Threshold = -1.1c	Lags only	Moving	-6.051	-5.868	0.782

Note: Table 4 has same structure as Table 2.

Table 5: Coefficients for Selected UK Models

	Petrol		Diesel	
Regressor	Coeff.	Std. Error	Coeff.	Std. Error
Constant	0.002	0.001	0.002	0.001
ecm^2_{t-1}	-0.233	0.090	-0.266	0.173
ecm_{t-1}	-0.434	0.120		
ecm^3_{t-1}	-0.493	0.117	-0.387	0.119
ΔPR^2_t			0.405	0.128
ΔPR^2_{t-1}			0.358	0.112
ΔPR_t	0.482	0.034	0.437	0.048
ΔPR_{t-1}	0.253	0.069		
ΔPR^3_t			0.630	0.089
ΔPR^3_{t-1}			0.287	0.103
ΔPC_{t-1}	0.135	0.062		

Note: The structure of Table 5 is similar to Table 3. For the petrol model, there is no asymmetry in the lags of the refined price so the superscripts do not apply.

Table 6: Comparison of Symmetric, 2 Regime and 3 Regime Models

PETROL IRELAND

Model	AIC	SBC	R^2
Symmetric	3.479	3.603	0.787
3 Regime Model	3.364	3.508	0.811
2 Regime Model	3.452	3.658	0.798

DIESEL IRELAND

Model	AIC	SBC	R^2
Symmetric	3.644	3.748	0.714
3 Regime Model	3.532	3.655	0.741
2 Regime Model	3.638	3.763	0.718

PETROL UK

Model	AIC	SBC	R^2
Symmetric	-6.377	-6.275	0.860
3 Regime Model	-6.382	-6.239	0.862
2 Regime Model	-6.403	-6.281	0.864

DIESEL UK

Model	AIC	SBC	R^2
Symmetric	-6.014	-5.912	0.768
3 Regime Model	-6.076	-5.914	0.786
2 Regime Model	-6.005	-5.863	0.769

Note: The 3 Regime Model is the model with the best performance statistics found in the earlier part of the paper. The 2 Regime model is chosen using the same criteria as the 3 regime model but regimes are defined according to the treshold variable.

4.10 Glossary of Terms used to Describe Models

There are a wide variety of models used in the paper. This glossary explains the terminology used and provides a brief description of the models. The terms used here appear in Tables 2, 4 and 6.

Symmetric: This is the standard short-run equation from a cointegration model.

Asymmetric ECM: In this type of model, the speed of adjustment parameters in the short-run equation are allowed to vary depending on whether the equilibrium error is positive or negative.

Zero Threshold Model: In this type of model, parameters in the short-run equation are allowed to vary depending on whether the threshold variable is above or below zero. Tables 2 and 4 outline the parameters which are permitted to vary and the threshold variable. The zero threshold models also divide the sample according to positive and negative equilibrium errors. This results in four different regimes but the two regimes with conflicting price pressures are merged to give a three regime model.

Non-Zero Threshold Model: This type of model allows the threshold variable to have a non-zero threshold value. In all other respects, it is identical to the zero threshold model.

2 Regime Model: This type of model allows parameters in the short-run equation to vary depending on whether the threshold variable is above or below its threshold value. The threshold value may be zero or non-zero. The value of the equilibrium error is irrelevant.

4.11 Appendix 1: Descriptive Statistics for Data

This appendix provides a more detailed breakdown of the data series used in the analysis. The main data used in the study are monthly data series on the refined prices of gas and diesel and pre-tax prices of petrol and diesel. Estimates are conducted over the period Jan '97 - April '09 for Ireland and Jan '97 - May '09 in the case of the UK. Some observations are lost at the beginning of each sample due to differencing and the inclusion of lags in the regressions. Summary statisics for the series are presented here:

Summary Statistics for Monthly Data Series

	MEAN	STD ERROR	MINIMUM	MAXIMUM
IRELAND				
Diesel Price	44.33	13.36	25.70	88.20
Change in Price	0.09	2.71	-10.51	8.20
Refined Diesel Price	25.35	13.23	6.97	66.48
Change in Refined Diesel	0.09	2.57	-10.70	9.17
Petrol Price	39.92	11.05	23.38	71.22
Change in Price	0.10	2.91	-12.04	10.78
Refined Petrol Price	23.96	11.19	6.41	54.67
Change in Refined Petrol	0.10	2.84	-12.20	5.37
UK				
Diesel Price	35.47	13.50	13.94	77.84
Change in Price	0.12	2.42	-10.77	7.63
Petrol Price	30.07	11.33	13.94	63.42
Change in Price	0.09	2.60	-12.27	5.14

Note: Prices are reported in euro cents. Refined prices are international prices. Statistics are calculated from January 1997 to mid 2009.

The statistics for the mean prices of refined products versus the mean pre-tax consumer prices for the UK and Ireland demonstrate that the gap between refined prices and consumer prices are higher on averge in Ireland. The standard error for the change in prices is very similar for both refined price and consumer prices. The mean price change understates the average magnitude of price changes as positive and negative changes tend to offset each other. For example, the average change for refined diesel is 0.09 but the average of the absolute change, which measures the average size of price changes, is 1.75c. The average absolute price change for UK consumer diesel prices is 1.58c. There is no obvious seasonal pattern in the data for the UK and Ireland. Correlation analysis of consumer and refined prices shows no evidence of seasonality. In addition, the growth rates of these prices show no tendency for unusally large or small values to occur in any specific month.

Summary Statistics for UK Weekly Data Series

	MEAN	STD ERROR	MINIMUM	MAXIMUM
Diesel Price	35.51	13.53	13.28	78.97
Change in Price	0.04	0.9	-6.1	4.66
Refined Diesel Price	25.37	13.2	6.82	68.47
Change in Refined Diesel	0.02	1.14	-5.44	4.66
Petrol Price	30.97	11.7	13.54	64.52
Change in Price	0.04	0.88	-6.32	3.91
Refined Petrol Price	23.08	11.27	6.27	56.44
Change in Refined Petrol	0.03	1.25	-6.24	7.89

Note: Statistics are calculated from January 3, 1997 to May 29, 2009.

The weekly statistics follow a similar pattern to the monthly. The main difference is with the series measuring the change in prices. The average change in prices is smaller when weekly data are examined and the standard error of these changes is also smaller relative to monthly. Once again, the correlation structure of the price series indicates that seasonality is not a feature of the data.

Chapter 5

Understanding and Forecasting Aggregate and Disaggregate Price Index Dynamics

The issue of forecast aggregation is to determine whether it is better to forecast a series directly or instead construct forecasts of its components and then sum these component forecasts. Notwithstanding some underlying theoretical results, it is generally accepted that forecast aggregation is an empirical issue. Empirical results in the literature often go unexplained. This leaves forecasters in the dark when confronted with the option of forecast aggregation. We take our empirical exercise a step further by considering the underlying issues in more detail. We analyse two price datasets, one for the United States and one for the Euro Area, which have distinctive dynamics and provide a guide to model choice. We also consider multiple levels of aggregation for each dataset. The models include an autoregressive model, a factor augmented autoregressive model, a large Bayesian VAR and a time-varying model with stochastic volatility. We find that once the appropriate model has been found, forecast aggregation can significantly improve forecast performance. These results are robust to the choice of data transformation.

5.1 Introduction

When forecasting economic variables, one is often faced with the choice of either forecasting an aggregate directly or forecasting its components and then summing the component forecasts. This is frequently encountered when forecasting inflation, where prices are commonly available for a large number of components series in addition to the aggregate price index. The aggregation issue is a major practical consideration when it comes to forecasting key economic indicators but frequently forecasters are in the dark in terms of which approach is likely to yield the best results. There is a considerable set-up cost when estimating models on disaggregate data if there are a large number of component series so researchers are understandably reluctant to pursue this strategy unless it is likely to yield benefits.

Arguably, the literature on forecast aggregation is at an impasse. The early contributions focussed on deriving theoretical results but this approach was eventually abandoned as the underlying assumptions were too restrictive. Empirical papers tend to focus on a specific application. Competing sets of forecasts are constructed for a given country or set of countries to see whether forecast aggregation helps. With the exception of Hubrich (2003), few papers offer potential explanations of why the forecast aggregation strategy was a success or failure so there is little guidance to forecasters faced with the option of combining disaggregate forecasts.

We conduct empirical exercises but relate our findings back to the properties of the dataset and the models used. The exercise is conducted on both United States (US) and Euro Area (EA) inflation. Although both datasets relate to inflation, these datasets have distinct characteristics and we tailor the model to the properties of the data. In contrast to most previous studies, we consider multiple levels of aggregation for each dataset.[1] We find that, once the appropriate model is found for a dataset, forecast aggregation always leads to improvements in forecast accuracy - the critical issue is to find the appropriate model. Frequently, the forecast based on the aggregate results in the worst forecast performance. This story is consistent with the theoretical literature. By providing a detailed explanation for main factors driving results for both datasets, we provide a greater understanding of the key issues relative to other empirical papers.

In the next section, we provide a summary of the main contributions in both the theoretical and empirical side of the literature. Section 3 briefly considers the issue of aggregation of forecasts in terms of an AR model. We do this from an heuristic perspective with a view to highlighting aggregation is likely to be beneficial when the series have contrasting dynamics and offsetting errors. Section 4 describes the data. Section 5 describes the models used in the paper with the results reported in section 6. Section 7 outlines some robustness checks while section 8 provides a summary and concludes the paper.

[1]One exception is Duarte and Rua (2005), who consider a 5-item and 59-item breakdown of the CPI in Portugal. We examine four different levels of aggregation for the US and three for the EA in this study.

5.2 Literature Review

Early contributions in the area of forecast aggregation were mainly confined to theoretical results based on an assumed data generating process (DGP). Assuming that the components are ARIMA processes, Rose (1977) examines the DGP and forecasts for an aggregate of these models. Others including Tiao and Guttman (1980), Kohn (1982) and Lutkepohl (1984a, 1984b) followed this approach with the DGP or forecast performance of the aggregated process related to an assumed structure for the DGP of the components. Based on asymptotic theory, it is possible to state that the disaggregate forecast will have a lower forecast error if the DGPs of all components are known. Lutkepohl (1984a) acknowledges that the superiority of the disaggregate forecast is no longer assured if the DGPs aren't known and instead must be estimated. In practice, DGPs are not known to forecasters so the results of these studies have limited practical implications.

European Monetary Union (EMU) revived interest in the topic of forecast aggregation but given the limited success of the theoretical approach, the literature changed direction and empirical exercises became much more common. There have been two distinct approaches adopted. The traditional approach, which is followed in this paper, is to construct forecasts of the disaggregates and combine them. In a couple of recent papers, Hendry and Hubrich (2006, 2010) suggest the alternative route of including disaggregates directly in the model of the aggregate. These two papers consider both predictability in population and forecastability in sample through both analytical and empirical work. They consider a number of practical issues such as changing coefficients, specification error and estimation uncertainty. They find that including disaggregate information in the aggregate models helps to improve forecasts. The work represents an alternative approach to improve forecasts through the use of disaggregate information and supports the broad concept of forecast aggregation.

This paper is concerned with the traditional approach, which is the focus of most of the literature. Hubrich (2003) and Benalal et al (2004) both examine HICP inflation for the euro area (Benanal et al also consider the four largest coutries) and find that there are no significant benefits to forecast aggregation. In country specific studies of HICP inflation, Duarte and Rua (2005), Bruneau et al (2007) and Moser et al (2007) all find forecast aggregation leads to improved forecasts for inflation for Portugal, France and Austria respectively. The results for the EA papers contrast with the country specific studies but all papers employ different models and are estimated over different time spans.

Forecast aggregation has also been examined in the context of output forecasting. Zellner and Tobias (2000) forecast the aggregate growth rate of 18 industrial countries using an aggregate and disaggregate approach. They report improved forecasts from the disaggregate approach. Marcellino, Stock and Watson (2003) forecast prices and three activity measures for the euro area directly and by aggregating country specific models. They find forecasts are more accurate when country specific models are aggregated. With the exception of Hubrich (2003) and Benalal et al (2004), the results of the empirical papers generally support forecasting disaggregates. Hubrich (2003) and Benalal et al (2004) are both reliant on short spans of data, as they were conducted shortly after the beginning of monetary union. This suggests that estimation error may have been a significant problem, particularly in the case of highly parameterised models such as VARs. They also focus on a small number of disaggregates - five in each case. We find that it is preferable to use a more detailed breakdown and our results support forecast aggregation for EA inflation.

The aim of this paper is to look into the issue of forecast aggregation in greater detail than the existing literature. In contrast to the standard approach, we utilise multiple levels of disaggregate data for each dataset. This allows us to explore the properties of the data which lead to benefits in terms of forecast aggregation. By considering two separate dataset with different characteristics, we are also able to highlight the importance that the selection of the correct model type has on the results. These insights are valuable to other forecasters contemplating the aggregation approach.

5.3 Factors Affecting Forecast Performance

In this section, we discuss the factors that are likely to impact on forecast performance. We do not provide conclusive theoretical results which determine the results of our empirical exercise. The theoretical models are not sufficiently rich to capture the interplay of all relevant factors in a unified framework. But a discussion of some of the relevant issues here helps to provide a more intuitive understanding of the empirical work. We frame the discussion in this section around the AR model as this is the most basic model that we use in the empirical exercise.

5.3.1 Specification of Aggregate Process

In the case examined in the paper, the price aggregate is a weighted average of all the other sub-components, and a as consequence, its dynamic are likely to be quite complex. For example, theoretical results tell us that the aggregate of two AR(1) processes will be an ARMA(2,1) process. More generally, the aggregate of an AR(p_1) and AR(p_2) process will be an ARMA(($p_1 + p_2$),max(p_1, p_2)) process. Thus, when aggregating a large number, say n, of component AR process, the theoretical AR lag length is $\sum_{i=1}^{n} p_i$, which may be even greater than the number of data points available when n is large. The theoretical MA lag is simply the longest MA lag found among the individual series. The estimated aggregate process will represent an approximation to this theoretical model, with a lot of the theoretical coefficients set equal to zero. All theortically relevant coefficients will not be statistically significant so some will be excluded in practice. The exclusion of relevant parameters is balanced against the need for parsimony. Amongst others, Enders (2010) points out that forecasts may be better from overly parsimonious models relative to those that exactly fit the theoretical model, as the former may benefit from low estimation error / parameter uncertainty. Therefore, in forecasting work, we might not worry about estimating all coefficients as long as the key parameters are tightly estimated.

5.3.2 Forecast Variance

If we again consider an AR process, the variance of the forecast error is known to depend on both the variance of the disturbance term and the amount of estimation error. Suppose that $y_t = ay_{t-1} + \epsilon_t$. In the absence of parameter uncertainty, the one-step forecast is given by $E_t y_{t+1} = ay_t$ with the corresponding mean squared forecast error (MSFE) given by $E_t(y_{t+1} - ay_t)^2 = E_t\epsilon_{t+1}^2 = \sigma_\epsilon^2$. To take account of parameter uncertainty, the known quantity a is replaced with $\hat{a}$ in the calculation of the MSFE:

$$\begin{aligned} MSFE &= E_t(y_{t+1} - \hat{a}y_t)^2 \\ &= E_t[(ay_t - \hat{a}y_t)^2) + \epsilon_{t+1}^2] \\ &= E_t[(a-\hat{a})^2](y_t)^2 + \sigma_\epsilon^2 \end{aligned} \tag{5.3.1}$$

with the later equalities holding due to independence assumptions. Clearly, the parameter estimation error is strictly positive and contributes to overall forecast variance. It is decreasing in terms of the sample size (and disappears asymptotically) but increasing in terms of the number of parameters. Thus, in a model with a short data span and a lot of parameters, it can prove a significant obstacle to forecasting. This underpins the need for parsimony in forecasting.

A serious shortcoming of the AR model is that a univariate specification takes no account of missing information. Prices can be influenced by a range of factors so the model will be miss-specified through omitted variable basis. The resulting coefficient bias will hurt forecast performance. The influence of the missing factors will show up in the residuals of each equation so that the fit of the model will be lower than desired. In the context of aggregating forecasts, there is a second implication. One would hope that the forecast errors of the individual series will offset each other to some extent. If, however, the missing information impacts the individual series in the same direction[2], then all the individual forecasts will tend to either overshoot or undershoot the correct value. This could seriously disadvantage the forecast aggregation approach. The key to overcoming omitted variable bias is by including extra information in the individual regressions in as parsimonious a manner as possible. We will return to comment on these issues in the results section.

[2]An positive oil price shock would cause most of the inflation series to increase.

5.4 Data

5.4.1 US Data

The analysis in this paper draws on both US data and EA data. The US series are NIPA data from the Bureau of Economic Analysis (BEA).[3] The price series are personal consumption expenditures available quarterly from 1959Q1 - 2009Q4. The data are already available at different levels of aggregation. This paper considers four different levels of aggregation for the dataset. The first is a three item breakdown which includes the prices of durable goods, non-durable goods and services. We next consider a fifteen item breakdown. The price categories are still quite broad at this level of aggregation and examples include food, housing and transport. A full list of price series for all levels of aggregation is provided in the Table 1. The third breakdown consists of fifty different price series The categories here are quite narrowly defined and again are presented in Table 1. The final breakdown is based on 169 series. The series are too numerous to list in the Table but a list of included items is available upon request.

As we wish to compare aggregated individual forecasts with the forecasts from the overall PCE inflation rate, we must be able to construct the PCE inflation rate from the individual inflation rates as a first step. This requires the weights of each item for each level of aggregation. All data are taken from Tables 2.4.4U, 2.4.5U and 2.4.6U on the BEA website. The price series are chained index values and their weights are calculated according to the approximation provided in Dolmas (2006):

$$w_{i,t+1} = \frac{1}{2}\frac{Q_{i,t}P_{i,t}}{\sum Q_{i,t}P_{i,t}} + \frac{1}{2}\frac{Q_{i,t+1}P_{i,t}}{\sum Q_{i,t+1}P_{i,t}}$$

The weight at time $t+1$ is equal to an average of the expenditure share of the product at time t and its expenditure share had consumers bought the $t+1$ quantity at time t prices. In each case, the accuracy of this approximation was checked by constructing the aggregate inflation rate from the components. The aggregate inflation rate was recovered with a high level of precision, which ensure the validity of the empirical exercise.

Figure 1 graphs the Year on Year (YoY) PCE inflation rate and its component inflation rates for each of the four different levels of aggregation used in the paper. In each graph, the thick blue line is the aggregate inflation rate. For the graph of the 3 items, the individual items move in tandem with the PCE inflation rate. As the number of items in each breakdown increases, the series obviously have more individual dynamics although there is still quite noticeable comovement with the PCE rate, indicated by the tight bunching of series around the PCE inflation rate.

[3] Available at: http://www.bea.gov/national/nipaweb/SelectTable.asp

5.4.2 EA Data

The euro area data are price series for the Harmonized Index of Consumer Prices (HICP). The series along with their weights are available on the Eurostat website.[4] The series are disaggregated at three different levels with a 5-item, 12-item and 32-item breakdown. The items included in each level of aggregation are also presented in Table 1. The series are monthly and the sample period is from January 1996 to December 2009. Although it is possible to get data at a more detailed level over the latter part of the sample, it is not possible to do so for the entire sample so the 32-item breakdown represents the most detailed available for our purposes. There is a strong seasonal pattern in some of the euro area data when month-on-month growth rates are calculated. Seasonally adjusted data are not available. In addition, the seasonal pattern is not stable over the sample and so it not possible to estimate a consistent seasonally adjusted series. To mitigate this problem, estimation is conducted using year-on-year growth rates. Seasonality is not an issue with the US data as all series are seasonally adjusted.

Figure 2 graphs the YoY HICP inflation rate and the component rates at the three levels of aggregation used. At the five item level, the series display more heterogeneous dynamics relative to the US data. This pattern is repeated with the 12 and 32 item datasets. Although there is bunching around the aggregate, these series have stronger individual characteristics than the US data. This is probably due to the fact that the US data are from one country whereas the EA data combines the inflation rates of several difference countries.

5.5 Data Transformation and Models Specification

The empirical exercise in this paper is addressed in the following way: we construct one set of forecasts by estimating models on the aggregate series and a second set by using the same model to forecast the individual series prior to aggregation, then we compare the accuracy of both approaches. The target variable is the aggregate, annualized h period inflation, defined as $\pi_t^h = k \log(\frac{P_t}{P_{t-h}})$, where the constant k is the normalization term.[5] P_t is the aggregate level of price index. Given a model m, we perform a pseudo out-of-sample forecasting simulation. At time t, we estimate the parameters of the model and compute the forecasts of the aggregate and disaggregate inflation series at horizon h, then we update the sample with a new observation and, at time $t+1$, we re-estimate the parameters of the model and compute again the forecasts for time $t+1+h$. The exercise is iterated up to the end of the sample.

Forecasts of the target variable at horizon h are denoted as $\hat{\pi}_{a,t+h|t}^{h,m}$, when they are computed directly on the aggregate inflation series, and as $\hat{\pi}_{d,t+h|t}^{h,m} = \sum_{j=1}^{N_s} w_{j,t}\hat{\pi}_{j,t+h|t}^{h,m}$

[4] Available at http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

[5] It is $\frac{400}{h}$ in the case of quarterly data and $\frac{1200}{h}$ in the case of monthly data.

when they are computed by aggregating forecasts of disaggregate inflation series j.[6] The first subscripts a or d denote if the forecast of the target variables is computed with the aggregate or disaggregate inflation series respectively, while $t+h|t$ refers to the fact that, for horizon h, forecasts are computed by using information up to time t. Finally, the first superscript h denotes the transformation adopted for prices, while m refers to the model employed.

Atkeson Ohanian Model (AO) (2001)

$$\pi^h_{t+h} = \pi^4_t + \omega^h_{t+h} \tag{5.5.2}$$

The forecast at $t+h$ is computed as:

$$\hat{\pi}^{h,AO}_{a,t+h|t} = \pi^4_t \tag{5.5.3}$$

AutoRegressive Model (AR)

For a generic inflation series j:

$$\pi^h_{j,t+h} = \alpha^h_j + B^h_j(L)\pi^h_{j,t} + \varepsilon^h_{j,t+h} \tag{5.5.4}$$

where $B^h_j(L) = B^h_{j,0} + ... + B^h_{j,s}L^s$ is a polynomial in the lag operator L. Parameters are estimated by Ordinary Least Squares (OLS). The forecast for a given horizon h is computed as:

$$\hat{\pi}^{h,AR}_{j,t+h|t} = \hat{\alpha}^h_j + \hat{B}^h_j(L)\pi^h_{j,t} \tag{5.5.5}$$

Factor Augmented AutoRegressive Model (FAAR)

$$\pi^h_{j,t+h} = \nu^h_j + C^h_j(L)\pi^h_{j,t} + \gamma F_t + \zeta^h_{j,t+h} \tag{5.5.6}$$

This is the AR model of *eq.*(5.5.4) augmented with one factor. The factor is estimated with the first principal component (Stock and Watson, 2002) computed on the most detailed data set available. For example, the factor for the US dataset is computed on the dataset of 169 series, while that for the Euro area is computed on the dataset of 32 series. The factor is added as an explanatory variable each of the disaggregate equations. The factor will enter with a different coefficient for each equation, depending on the relationship between the factor and the individual series. The working of the factor model is examined in more detail when results are discussed in section 5.6.2. Parameters of *eq.*(5.5.6) are estimated by OLS. The forecast at horizon h is given by:

[6] N_s is the number of series in the sth set of disaggregate series; $s = \{1, 2, 3, 4\}$ in the case of US dataset and $s = \{1, 2, 3\}$ in the case of Euro area dataset.

$$\hat{\pi}^{h,FAAR}_{j,t+h|t} = \hat{\nu}^h_j + \hat{C}^h_j(L)\pi^h_{j,t} + \hat{\gamma}\hat{F}_t \tag{5.5.7}$$

Bayesian VAR (BVAR)

This is a Bayesian VAR with the Minnesota prior as proposed by Banbura, Giannone and Reichlin (2010). Let's denote with $P_{j,t}$, the price level for series $j \in S_i, S_i = \{1, ..., j, ..., n_i\}$; the model is estimated on the log-level of the series denoted as $p_{S_i,t}$:

$$p_{S_i,t} = c + A_1 p_{S_i,t-1} + ... + A_p p_{S_i}, t - p + v_t \tag{5.5.8}$$

where $p_{S_i,t}$ is a $(S_i \times 1)$ vector of variables, c is a $(S_i \times 1)$ vector of constants, $A_1...A_p$ are $(S_i \times S_i)$ matrices of coefficients and v_t is a $(S_i \times 1)$ vector of disturbances. The estimation of the model for a large set of variables is unfeasible due to the curse of dimensionality.

One solution is to impose restrictions (prior beliefs) on the parameters of the system. Following Banbura, Giannone and Reichlin (2010) we impose Litterman (1986) priors. The coefficients of a matrix A_i, $i = 1, ..., p$ are normally distributed random variables with the mean of the coefficient matrix on the first lag (matrix A_1) equal to an identity matrix I_{S_i} and the mean of all the other coefficients equal to zero. The variance of the parameters depends on a parameter τ which defines the tightness of the priors. A value of τ equal to zero exactly imposes the random walk with drift model on the variables, while a value of τ bigger than zero allows for some variability around the mean of the coefficients and the random walk prior is not exactly imposed.[7] We impose also another type of prior, on the sum of coefficients of the matrices $A_1...A_p$. This prior is imposed by means of another parameter μ. If $A_1 + ... + A_p = I_{n_i}$ the prior is imposed exactly and the specification is equivalent to a VAR in first differences. This will imply that the forecasts will converge to the variable's growth rate. A forecast of the log level of series j at horizon h is then computed as:

$$\hat{p}_{S_i,t+h|t} = \hat{c} + \hat{A}_1 \hat{p}_{S_i,t+h-1|t} + ... + \hat{A}_p \hat{p}_{S_i,t+h-p|t} \tag{5.5.9}$$

where $\hat{p}_{S_i,t+h-i|t} = p_{S_i,t+h-i}$ if $i >= h$. In practice, the forecast at time $t+h$ is computed recursively from the forecast at time $t+1$. The estimates of the parameters correspond to the median of the posterior distributions. For each series $j \in S_i$, the annualized h period inflation rate is computed as:

$$\hat{\pi}^{h,BVAR}_{j,t+h|t} = (\hat{p}_{j,t+h|t} - p_{j,t})k \tag{5.5.10}$$

The maximum lag length for the AR model is specified ex ante and the actual lag length is chosen according to the Bayes Information Criterion (BIC), while the lag specification for the BVAR is selected by choosing the number of lags that minimize the squared forecast errors of the previous period, including a minimum of 5 lags for

[7] A scale parameter, to fix the variance of the coefficients, is set by estimating the variance of the residuals from a univariate model of order p on the single variables $p_{j,t}$.

the quarterly dataset and a minimum of 13 lags for the monthly dataset. The values of the hyperparamters, μ and τ are chosen on a grid search so that the fitted model has an R^2 as close as possible to 50%. This ensures a reasonable in-sample fit but guards against over-fitting, which leads to poor out-of-sample forecasts.

The first estimation sample (prior to data transformations) for US data is 1959:Q1-1995:Q2 with forecasts beginning from 1995Q2+1 to 1995Q2+h. This is the earliest sample for which weights are available. Final recursive forecasts are computed up to 2009Q4. For the EA data, the first estimation sample is 1996:M1-2001:M12 and the forecast begins in 2000M12+1 to 2000M12+h. The final recursive forecasts end in 2009M12. Forecast accuracy is evaluated through the Mean Square Forecast Error (MSFE) statistic, however, to facilitate the comparison, the accuracy of a model m is compared (ratio) with that obtained by the Atkeson-Ohanian random walk model, used as the benchmark. Finally, as already mentioned above, some series in the EA dataset are characterized by strong seasonal patterns. To mitigate seasonality the exercise is performed on the year-on-year data transformation. *Eq.*(5.5.4) for example is modified as follows:

$$\pi^4_{j,t+h|t} = \alpha^4_j + B^4_j(L)\pi^4_{j,t} + \varepsilon^4_{j,t+h}$$

where the superscript 4 refers to the data transformation.[8]

5.6 Results

5.6.1 AR

The numbers in Table 2 are ratios of the RMSE from the AR model relative to the AO benchmark, with a value less than one indicating the forecast model outperforms the benchmark at the specified horizon. The first part of the table shows the results for the United States. When an AR model is used to forecast the aggregate directly, it is only possible to improve upon the benchmark at the one-quarter horizon. Forecast performance relative to the benchmark gets relatively worse as the horizon increases. When AR models are applied to the disaggregates and combined, the forecasts are improved. The AR model based on the 15-item aggregation provides the best forecasts, as they outperform the benchmark up to 5 quarters ahead. This information is also presented in graphical format in Figure 3. It clearly shows how poorly the aggregate forecast performs relative to the disaggregates over most of the horizon. Another feature evident from the graph is that the greatest gains in forecast accuracy are given by the 3-item and 15-item breakdown. Forecast accuracy decreases, particularly over longer horizons, when the 50-item and 169-item breakdown are used. Considering the US AR model in isolation, any gains in forecasting the disaggregates have already been exploited when the 15-item breakdown is used.

[8] As robustness check, in the last section of the paper, we report a set of results for an alternative price transformation (for the US dataset).

The second part of Table 2 presents the results for the Euro Area. The horizon is now twelve months rather than eight quarters and there are only three levels of aggregation rather than four. The results here differ from the US results in a number of ways. The forecasts here are better than the US forecasts in the sense that improvements relative to the benchmark are much greater. Furthermore, the greatest gains relative to the benchmark are at the longer horizons for the Euro Area whereas the US model forecasts have their greatest gains at the shorter horizons. This behaviour is largely explained by the performance of the benchmark. The HICP is not as persistent as the PCE inflation rate. Consequently, the AO benchmark is not as good for the Euro Area, particularly at longer horizons i.e. the benchmark is much easier to beat at longer horizons in the Euro Area due to the higher mean reversion of this series. This is why the graphs slope downwards although, in absolute terms, the forecast errors are still increasing in line with the forecast horizon. The key messages are the same as for the US however. The results in the table again show that the aggregate performs poorly relative to the disaggregates. The 32-item breakdown results in the best forecasts. The results are also graphed in Figure 4 and the difference in the performance of the aggregate relative to the disaggregates is quite stark.

5.6.2 FAAR

The first section of Table 3 documents the forecast performance for the US when a factor is included in the forecast equation. The aggregate now improves upon the benchmark when forecasting up to three quarters in the future. The aggregate forecast still has the least satisfactory performance however, which is also evident graphically in Figure 5. The model based on 169 disaggregates now has the best forecast performance over most horizons. The factor forecasts outperform the simple AR model so that the best forecasts overall for US inflation come from the 169-item model. The EA results with the FAAR are remarkably similar to those for the AR model. The aggregate forecast is strongly outperformed by the disaggregates, with the 32-item breakdown yielding the best results and this pattern is clearly evident in Figure 6. The results of Tables 2 and 3 are supportive of forecast aggregation, with the most accurate forecasts coming from disaggregate models.

Table 4 documents the change in forecast accuracy when the factor is added to the forecasting equation. There are universal improvements in forecast accuracy for the US data. In addition, the forecasts based on the most detailed breakdown enjoy the greatest improvements in forecast performance. In contrast, for the EA models, the inclusion of the factor leads to virtually no change in the forecasts. To examine the reasons for this, we examine the structure of the dataset to see if there is a large common element to the series. Firstly, we regress the individual series on the factor alone and report the average R^2 in Table 5. We also calculate the average correlation between the series. There is strong commonality in the US PCE dataset. One factor explains 85% of the variation in the aggregate series. The average R^2

declines in line with the number of disaggregates but even at the 169-item level, the average R^2 is 30%. Similarly, the average correlation between the series is high at this level. Commonality is much lower for the euro area inflation series. One factor only explains 37% of the variation in the aggregate series. This drops to about 20% for the disaggregates. Similarly, the average correlation between the series is low.

The strong common element in the US dataset is picked up by the factor model. The simple AR model which excludes the factor is, therefore, mis-specified via the omission of a relevant variable. This will have the usual effect of creating a bias in the coefficients, which will obviously impact the forecasts. Although not reported, the correlation amongst the residuals for the US AR models was found to be far higher than their EA counterparts, as the common factor was captured by the residuals and this imparted much stronger correlation. As described earlier, higher correlation amongst the disaggregate residuals is not good for forecasting. This underlying structure in the datasets explains why the factor needs to be included for the US models but not for the EA models.

The one outstanding issue is why the forecasts based on more detailed data improve to a greater extent. The aggregate and 3-item US FAAR models have very modest improvements in forecast power relative to their AR counterparts. The PCE aggregate and the 3-item inflation rates are a weighted average of a large number of underlying inflation rates. Similarly, the factor is a weighted average of all inflation rates. Although the factor weights are calculated to satisfy a maximum variance criterion, the aggregate inflation rate and the 3-item inflation rates are much like factors as they pick up a lot of the commonality in the data. Consequently, the aggregate and 3-item AR models are effectively modelling the common component. Thus, improvements over the AR model are relatively modest when the factor is added as the factor and AR series both model the same component of the data. At the more detailed levels of aggregation, the AR component can pick up the stronger individual dynamics while the factor picks up the common element, which is still meaningful even for the 169-item breakdown. The AR component and the factor are now modelling different behaviour. This is why forecast improvements are greater for the detailed breakdown. This demonstrates the interplay between model choice and the level of aggregation in the data.

5.6.3 BVAR

Table 6 presents the results of the BVAR model. The US results show that the BVAR is also a fruitful way to exploit the dynamic properties of the data, with the forecast errors again much smaller than those from the standard AR model, particularly for short horizon forecasts. In comparison with the factor model, the BVAR tends to perform better for the short horizons and the factor model does better over the longer horizons. By averaging first by horizon and then by level of aggregation, we find that the BVAR forecasts are equally accurate to the factor model forecasts. The exception to this pattern is the 15-item BVAR, which outperforms the 15-item factor forecast over all horizons. The best individual forecast of all forecast methods considered to date is also the 15-item BVAR, which is 6% more accurate than the next best model, when averaged by horizon. This is the only model which is more accurate than the benchmark at all horizons. Figure 7 graphs the performance of the BVAR models for the US and the superiority of the 15-item specification is clear. As the performance of the BVAR and the factor model are similar, one might tend to favour the factor model in the sense that it is easier to implement. In particular, the Bayesian approach relied on a grid search procedure to choose the hyperparameters in this application whereas the factor approach only relies on a decomposition of the covariance matrix of the inflation series.

The best BVAR model for the EA is the 5-item model, which is depicted in Figure 8. However, the results for the EA show that the BVAR fails to improve on the simple AR model. The strong individual dynamics of the series for the HICP means that the simple AR model provides the best forecast. Any attempt to capture common comovement or feedback between the series does not improve the forecasts. The BVAR forecasts are also weaker than the factor model, due to a drop in accuracy for the longer horizons but this is of less significance here as both are outperformed by the AR model.

5.7 Robustness Checks

5.7.1 Alternative Data Transformation

In this section we perform the forecasting exercise for a different price transformation, the h-level change in log prices $p_{t+h} - p_t$. This is analyzed only for the US dataset, given that the seasonal issues with the EA data makes difficult to look at alternative data transformations beyond the year-on-year inflation rate, which is analyzed in the previous section. At a given horizon h, the AR and FAR forecasting equations are exactly those of *eq.*(5.5.4) and *eq.*(5.5.5) respectively. Forecasts with the BVAR model are computed exactly as before, but the log-level of price at time t is then subtracted by the forecast of p_{t+h} in order to recover the h-level change of log-prices.

Table 7 presents the results for the standard AR model. The aggregate does not perform well, as it beats the benchmark only for the one-period forecast. The 3-item and 15-item forecasts are both more accurate than the benchmark up to five quarters. The results are presented graphically in Figure 9. The pattern mimics the year-on-year results, where the 3-item and 15-item forecasts are far better than the benchmark. Table 8 presents the results with the factor included. In further agreement with the year-on-year results, the aggregate still has the worst average performance and the 169-item model now has the most accurate forecasts. Figure 10 plots the results and it demonstrates how quickly the performance of the aggregate deteriorates over the forecast horizon from a strong starting position. Table 9 shows that the BVAR models also improve significantly on the AR specification, especially in relation to the long-range forecasts. The performance of the BVAR models is graphed in Figure 11. For the multistep forecasts in the previous section, the 15-item BVAR is the only breakdown which is clearly better than its factor comparator. For the forecasts considered here, the 3-item BVAR also outperforms the 3-item factor model due to good performance at the short horizons. Thus, when we compare the two model types averaged by all their forecasts, the BVAR is more accurate than the factor model by approximately 7%. As before, the best BVAR model is still based on 15 items. Overall, these results strongly mirror those of the year-on-year specification. In this exercise, the aggregate model never provides the best forecasts and often provides the worst. The key properties of the data affect the disaggregate forecasts in the same way irrespective of the data transformation.

5.7.2 Alternative Model for US Data

As a final robustness check, we consider one alternative model for the US. Our attention is limited to the US because this model constructs forecasts iteratively using quarter-on-quarter growth. The final type of model considered is a time-varying parameter AR model with stochastic volatility (TV-AR). D'Agostino et al (2009) estimate a TV-AR model for three macro variables in the U.S and find it does particularly well at forecasting inflation. The computational cost of estimating the TV-AR model means that it is only likely to be applied to a small number of items in practice. For this reason, we only conduct a partial exercise in which we estimate the model for the 15-item breakdown. The BVAR with 15 items is the most accurate model so it is instructive to use this as a comparator.

We assume that:

$$\pi^1_{j,t} = \delta_{j,t} + \rho_{1,t}\pi^1_{j,t-1} + ... + \rho^1_{p,t}\pi_{j,t-p} + e^1_{j,t} \tag{5.7.11}$$

where $\delta_{j,t}$ is the time varying intercept, $\rho_{i,t}$ with $i = 1, ..., p$ are time varying coefficients and $e^1_{j,t}$ is a Gaussian white noise with zero mean and time-varying variance σ^2_t. We assume that σ_t evolves as geometric random walk, belonging to the class of models known as stochastic volatility.

$$log(\sigma_t) = log(\sigma_{t-1}) + u_t \tag{5.7.12}$$

Forecasts at time $t + h$ are computed iteratively:

$$\hat{\pi}^1_{j,t+h|t} = \hat{\delta}_{j,t} + \hat{\rho}_{1,t}\hat{p}^1_{j,t+h-1} + ... + \hat{\rho}_{p,t}\hat{p}^1_{j,t+h-p} \tag{5.7.13}$$

where $\hat{\pi}^1_{j,t+h-i} = \pi^1_{j,t+h-i}$ if $i >= h$. The estimates of the parameters correspond to the median of the posterior distributions.[9]

[9] We fix $\lambda_1 = \lambda_2 = 10e^{-02}$. These are the tightness parameters governing the amount of time-variation in the coefficients and volatility respectively.

A technical issue arises when we generate multi-step expectations; we have to evaluate the future path of drifting parameters. We follow the literature and treat those parameters as if they had remained constant at the current level. See Sbordone and Cogley (2008) for a discussion of the implications of this simplifying assumption.

For each series j forecasts of are first cumulated to recover the h period inflation:

$$\hat{\pi}^{h,TV-AR}_{j,t+h|t} = \frac{1}{h}\sum_{s=1}^{h}\hat{\pi}^{1,TV-AR}_{j,t+s|t} \tag{5.7.14}$$

and are then aggregated to recover the forecast for the aggregate index:

$$\hat{\pi}^{h,TV-AR}_{d,t+h|t} = \sum_{j=1}^{N_s} w_{j,t}\hat{\pi}^{h,TV-AR}_{j,t+h|t} \tag{5.7.15}$$

The results for this exercise are presented in the first two columns of Table 10. As before, the first results column of the table shows the RMSE of the AO benchmark. The second column shows the forecasts errors of the TV-AR relative to the AO, with a value less than one indicating that the TV-AR has the better forecast. The third column compares the TV-AR to the BVAR. The results in the second column show that the forecast errors compare favourably to the benchmark over the entire forecast horizon. When compared to the BVAR, the TV-AR has more accurate forecasts over most horizons. The TV-AR does well for the short-term forecasts but its edge relative to the BVAR steadily declines to the point where the BVAR does better for quarters 7 and 8. Taken on average however, the TV-AR has the better performance with forecasts that are 6% more accurate on average over all horizons. The results demonstrate that combining forecast aggregation with time variation in the parameters and allowing for stochastic volatility can lead to even greater improvements in forecast performance. As the comparison for the TV-AR models is based on 15-items, we graph the results of the 15-item breakdown for all models in Figure 12. It shows that the AR model is not appropriate for a dataset with these properties. There is a big improvement moving to the factor model and further improvements when the BVAR and TV-AR models are used.

5.8 Summary and Conclusions

In this paper, we conduct an empirical exercise to test if it is possible to achieve gains in forecast accuracy by forecasting the individual components of inflation and aggregating the individual forecasts relative to forecasting the aggregate inflation rate directly. The empirical exercise uses data on both United States and Euro Area inflation. These datasets are quite distinct and require a different modelling approach. We consider four levels of disaggregation for the United States and three for the Euro Area. In all the empirical exercises in this paper, forecast aggregation leads to better forecasts. The aggregate forecast often has the least satisfactory performance and this makes the argument for aggregation more compelling given that multiple levels of aggregation are used.

The performance of the aggregated forecasts also depends on the type of model used. In particular, the model must capture the key characteristics of the data. There is strong comovement in US inflation. Simple AR models do not perform very well in this context but multivariate models such as factor models and BVAR models that can capture this common movement or pick up feedback between the series have more accurate forecasts. For the Euro Area inflation rate, there is far less commonality and the series have more individual dynamics. Simple AR models tend to work well for this type of dataset. They have more accurate forecasts than both the benchmark and their multivariate counterparts.

The exercises are mainly based on multistep forecasts of year-on-year inflation rates. For US inflation, we forecast the h-quarter price change for $h = 1..8$ and find the results are robust to this change in the target forecast variable. We also introduce a time-varying model with stochastic volatility where forecasts are constructed iteratively. The time-varying model in conjunction with forecast aggregation leads to further improvements in forecast power. These robustness checks corroborate the main results in favour of forecast aggregation. The paper provides a substantive endorsement of the forecast aggregation approach, particularly in terms of inflation. The key to realising gains in terms of forecast aggregation lies in the ability to uncover the appropriate model for a particular dataset.

5.9 Figures and Tables

Table 1: List of Items in each Aggregate

	HICP Inflation Aggregates		
5 Item List	**12 Item List**	**32 Item List**	
Processed Food	Food + beverages	Food	Health
Unprocessed Food	Alcohol + Tobacco	Non-alcoholic beverages	Purchase of vehicles
Non-Energy Goods	Clothing + Footwear	Alcoholic beverages	Vehicles operation
Energy	Housing	Tobacco	Transport services
Services	Furnishing	Clothing	Postal services
	Health	Footwear	Telephone and telefax
	Transport	Rents for Housing	Electronic Equipment
	Communications	Housing Maintenance	Other durables for recreation
	Recreation + culture	Water supply + misc. services	Recreation, garden and pets
	Education	Electricity, gas and fuels	Recreation services
	Restaurants + hotels	Furniture and furnishings	Reading and stationary
	Miscellaneous	Textiles	Holidays
		Appliances	Education
		Ware and Utensils	Catering services
		Tools and Equipment	Accommodation services
		Routine Maintenance	Miscellaneous
	PCE Inflation Aggregates		
3 Item List	**13 Item List**	**50 Item List**	
Durables	Motor vehicles and parts	New motor vehicles	Exp. abroad by US residents
Non-Durables	Durable household equipment	Used motor vehicles	Less remittances to nonresidents
Services	Rec. goods and vehicles	Vehicle parts	Housing
	Other durable goods	Furniture and furnishings	Household utilities
	Food and bev off-premises	Household appliances	Outpatient services
	Clothing and footwear	Household utensils	Hospital and nursing homes
	Gas and other energy goods	Equipment for house and garden	Motor vehicle services
	Other nondurable goods	Video, audio and IT equipment	Public transportation
	Housing and utilities	Sporting equipment	Parks, theaters,museums etc
	Health care	Sports and recreational vehicles	Audiovisual + IT services
	Transportation services	Recreational books	Gambling
	Recreation services	Musical instruments	Other recreational services
	Food service + accomm	Other durable goods	Food services
	Financial services	Food+ non-alc. bev. off-premises	Accommodations
	Other services	Alcoholic beverages off-premises	Financial services
		Food produced + consumed on farm	Insurance
		Garments	Communication
		Other clothes and footwear	Education services
		Gas + other energy goods	Professional and other services
		Pharmaceutical + medical products	Personal care and clothing services
		Recreational items	Social serv + religious activities
		Household supplies	Household maintenance
		Personal care products	Foreign travel by US Residents
		Tobacco	Less Exp in US by nonresidents
		Newspapers and magazines	Nonprofit Institution Exp.

Note: Some categories have been abbreviation. The list for the 169-item breakdown is available upon request.

Table 2: Forecast Errors for Standard AR Models

	United States					
Quarter	**AO**	**Aggregate**	**3 Items**	**15 Items**	**50 Items**	**164 Items**
1	0.35	0.74	0.66	0.69	0.67	0.67
2	0.73	1.02	0.86	0.81	0.84	0.86
3	1.16	1.08	0.91	0.85	0.90	0.91
4	1.63	1.22	1.01	0.94	1.01	1.02
5	1.69	1.28	1.07	0.95	1.08	1.08
6	1.56	1.41	1.20	1.09	1.29	1.29
7	1.48	1.53	1.32	1.26	1.50	1.52
8	1.32	1.62	1.49	1.50	1.79	1.84
	Euro Area					
Month	**AO**	**Aggregate**	**5 Items**	**12 Items**	**32 Items**	
1	0.07	0.92	0.76	0.78	0.75	
2	0.18	0.92	0.67	0.71	0.62	
3	0.33	0.99	0.62	0.69	0.58	
4	0.49	1.00	0.57	0.65	0.56	
5	0.68	0.98	0.56	0.64	0.56	
6	0.83	0.96	0.54	0.60	0.52	
7	0.99	0.91	0.54	0.57	0.50	
8	1.18	0.87	0.56	0.54	0.49	
9	1.34	0.81	0.57	0.54	0.50	
10	1.51	0.75	0.57	0.52	0.48	
11	1.67	0.68	0.57	0.48	0.47	
12	1.84	0.64	0.57	0.47	0.46	

Note: The table presents ratios of RMSE for each model relative to the benchmark. A value less than one indicates that the model has more accurate forecasts than the benchmark. The RMSE of the Atkeson-Ohanian benchmark is in the first column.

Table 3: Forecast Errors for FAAR Model

	United States					
Quarter	**AO**	**Aggregate**	**3 Items**	**15 Items**	**50 Items**	**164 Items**
1	3.09	0.66	0.63	0.66	0.64	0.60
2	2.27	0.85	0.79	0.76	0.75	0.72
3	1.82	0.93	0.87	0.79	0.78	0.74
4	1.63	1.09	0.96	0.86	0.85	0.81
5	1.42	1.16	0.98	0.83	0.84	0.77
6	0.98	1.32	1.07	0.92	0.95	0.87
7	0.77	1.48	1.19	1.03	1.09	1.02
8	0.62	1.59	1.33	1.18	1.27	1.23

	Euro Area				
Month	**AO**	**Aggregate**	**5 Items**	**12 Items**	**32 Items**
1	0.07	0.94	0.78	0.80	0.77
2	0.18	0.95	0.68	0.73	0.63
3	0.33	1.00	0.63	0.70	0.58
4	0.49	1.00	0.58	0.65	0.57
5	0.68	0.97	0.56	0.64	0.56
6	0.83	0.95	0.54	0.59	0.52
7	0.99	0.91	0.54	0.57	0.50
8	1.18	0.87	0.56	0.54	0.50
9	1.34	0.82	0.57	0.55	0.50
10	1.51	0.75	0.57	0.52	0.48
11	1.67	0.69	0.57	0.49	0.47
12	1.84	0.64	0.57	0.47	0.46

Note: The table presents ratios of RMSE for each model relative to the benchmark. A value less than one indicates that the model has more accurate forecasts than the benchmark.

Table 4: Change in RMSE by Including Factor

	United States				
Quarter	**Aggregate**	**3-item**	**15-item**	**50-item**	**169-item**
1	0.89	0.95	0.96	0.95	0.90
2	0.83	0.92	0.93	0.89	0.83
3	0.86	0.95	0.93	0.87	0.81
4	0.90	0.95	0.92	0.84	0.79
5	0.91	0.91	0.88	0.78	0.71
6	0.94	0.89	0.84	0.74	0.67
7	0.97	0.90	0.82	0.73	0.67
8	0.98	0.89	0.79	0.71	0.67
	Euro Area				
Month	**Aggregate**	**5 Items**	**12 Items**	**32 Items**	
1	1.02	1.02	1.02	1.03	
2	1.03	1.02	1.03	1.03	
3	1.01	1.02	1.01	1.01	
4	1.00	1.01	1.01	1.01	
5	0.99	1.01	1.00	1.00	
6	0.99	1.00	1.00	1.00	
7	1.00	1.00	1.00	1.00	
8	1.00	1.01	1.00	1.00	
9	1.01	1.00	1.00	1.00	
10	1.01	1.00	1.00	1.00	
11	1.01	1.01	1.01	1.01	
12	1.01	1.01	1.01	1.01	

Note: The table presents ratios of RMSE for AR models which include a factor to those that don't. It's a measure of the change in forecast accuracy as a result of including the factor in the model. A value less one means the model with the factor has more accurate forecasts.

Table 5: Commonality within Datasets

PCE	**Aggregate**	**3-item**	**15-item**	**50-item**	**169-item**
R^2	0.85	0.66	0.50	0.37	0.30
Ave. Corr.	n/a	0.73	0.61	0.47	0.39
HICP	**Aggregate**	**5-item**	**12-item**	**32-item**	
R^2	0.37	0.21	0.18	0.21	
Ave. Corr.	n/a	0.16	0.12	0.11	

Note: The individual inflation rates are regressed on the factor only. R^2 is the average for each level of aggregation. The second row is average correlation between the inflation rates.

Table 6: Forecast Errors for BVAR Model

	United States				
Quarter	AO	3-item	15-item	50-item	169-item
1	3.09	0.50	0.51	0.50	0.50
2	2.27	0.71	0.67	0.70	0.69
3	1.82	0.85	0.77	0.82	0.81
4	1.63	0.99	0.85	0.93	0.92
5	1.42	1.05	0.83	0.96	0.93
6	0.98	1.13	0.85	1.06	1.00
7	0.77	1.19	0.87	1.14	1.09
8	0.62	1.31	0.93	1.27	1.23
	Euro Area				
Month	AO	5 Items	12 Items	32 Items	
1	0.07	0.67	0.75	0.81	
2	0.18	0.55	0.58	0.61	
3	0.33	0.55	0.56	0.57	
4	0.49	0.54	0.55	0.57	
5	0.68	0.55	0.55	0.58	
6	0.83	0.56	0.56	0.59	
7	0.99	0.60	0.62	0.65	
8	1.18	0.62	0.65	0.70	
9	1.34	0.65	0.68	0.73	
10	1.51	0.67	0.69	0.75	
11	1.67	0.69	0.71	0.77	
12	1.84	0.72	0.72	0.80	

Note: The table presents the performance of the BVAR. As this is a multivariate model, there are no results to report for the aggregate alone.

Table 7: Errors for AR model Forecasts of $P_{t+h} - P_t$

	United States					
Quarter	AO	Aggregate	3 Items	15 Items	50 Items	164 Items
1	3.09	0.90	0.92	0.92	0.90	0.93
2	2.27	1.05	0.96	0.92	0.93	0.96
3	1.82	1.16	0.98	0.96	0.99	1.04
4	1.63	1.17	0.98	0.99	1.06	1.09
5	1.42	1.14	0.99	0.98	1.09	1.15
6	0.98	1.29	1.10	1.15	1.35	1.48
7	0.77	1.45	1.23	1.37	1.65	1.85
8	0.62	1.54	1.35	1.65	2.02	2.35

Note: The table presents ratios of RMSE for each model relative to the benchmark. A value less than one indicates that the model has more accurate forecasts than the benchmark. The RMSE of the Atkeson-Ohanian benchmark is in the first column. Only applies to US data.

Table 8: Errors for FAAR Model Forecasts of $P_{t+h} - P_t$

	United States					
Quarter	AO	Aggregate	3 Items	15 Items	50 Items	164 Items
1	3.09	0.85	0.93	0.91	0.88	0.89
2	2.27	0.97	0.96	0.90	0.87	0.84
3	1.82	1.09	1.00	0.92	0.90	0.85
4	1.63	1.11	1.01	0.93	0.91	0.84
5	1.42	1.09	1.01	0.88	0.88	0.81
6	0.98	1.28	1.09	0.91	0.94	0.86
7	0.77	1.46	1.19	1.00	1.05	0.97
8	0.62	1.54	1.27	1.09	1.17	1.10

Note: The table presents ratios of RMSE for each model relative to the benchmark. A value less than one indicates that the model has more accurate forecasts than the benchmark. Only applies to US data.

Table 9: Errors for BVAR Model Forecasts of $P_{t+h} - P_t$

	United States				
Quarter	**AO**	**3-item**	**15-item**	**50-item**	**169-item**
1	3.09	0.84	0.89	0.90	0.91
2	2.27	0.81	0.86	0.92	0.92
3	1.82	0.81	0.86	0.94	0.95
4	1.63	0.80	0.82	0.92	0.94
5	1.42	0.78	0.76	0.91	0.93
6	0.98	0.88	0.76	0.92	0.92
7	0.77	0.97	0.76	0.93	0.93
8	0.62	1.10	0.79	0.96	0.93

Note: The table presents the performance of the BVAR. As this is a multivariate model, there are no results to report for the aggregate alone.

Table 10: US Time-Varying AR Model Based on 15 Items

	$P_{t+h} - P_t$		
Quarter	**AO**	**TV-AR/AO**	**TV-AR/BVAR**
1	3.09	0.78	0.88
2	2.27	0.75	0.87
3	1.82	0.74	0.87
4	1.63	0.73	0.89
5	1.42	0.69	0.90
6	0.98	0.72	0.94
7	0.77	0.79	1.03
8	0.62	0.88	1.11

Note: The second results column shows the RMSE of the 15-item TV-AR while the third column is the ratio of the RMSE of the 15-item TV-AR to the 15-item BVAR.

Figure 1: PCE Inflation and its Component Inflation Rates

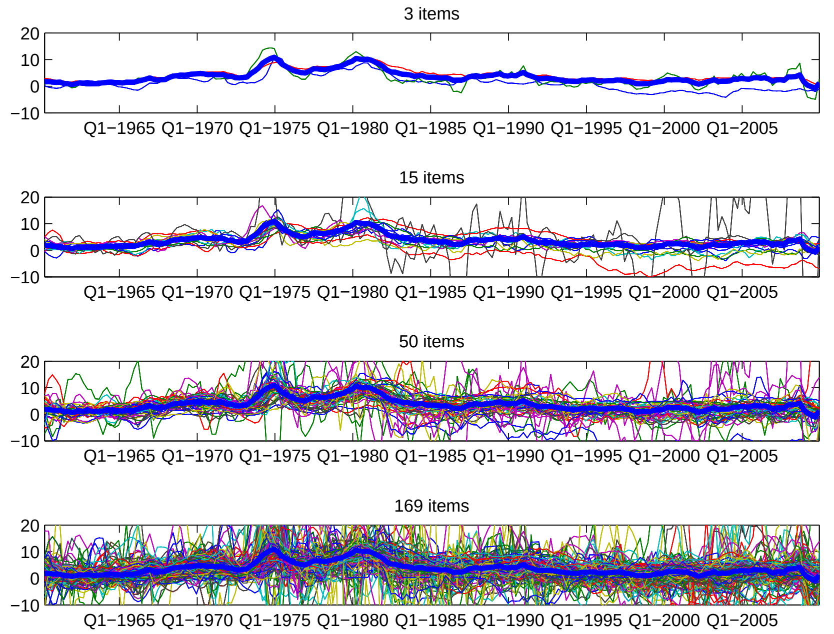


Figure 2: HICP Inflation and its Component Inflation Rates

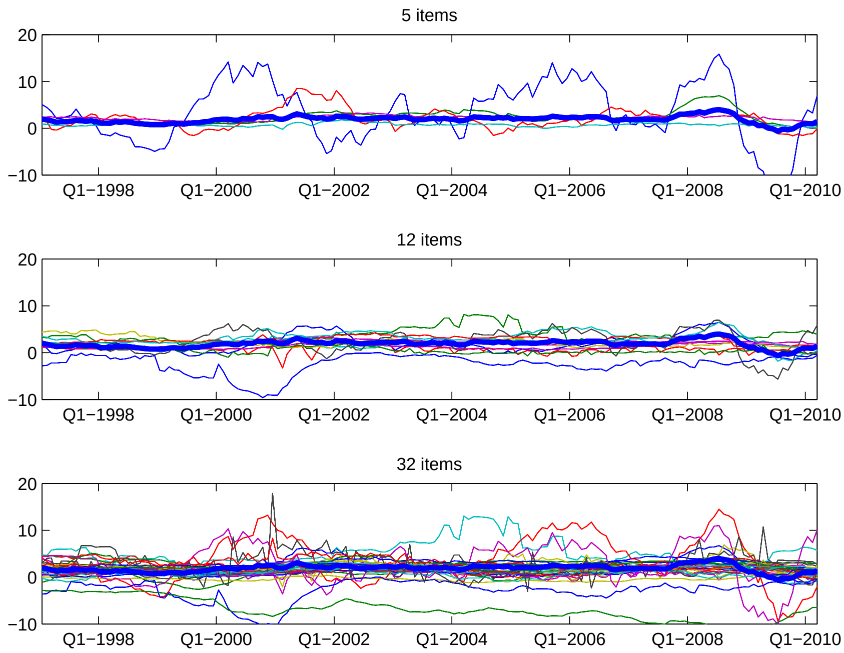


Figure 3: RMSFE of AR Models versus AO (US-YoY)

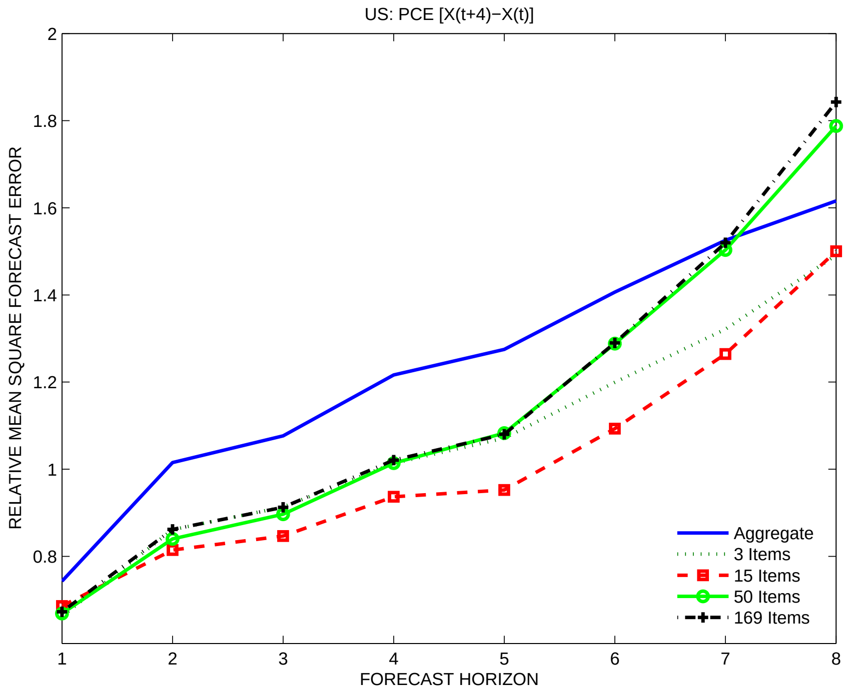


Figure 4: RMSFE of AR Models versus AO (EA-YoY)

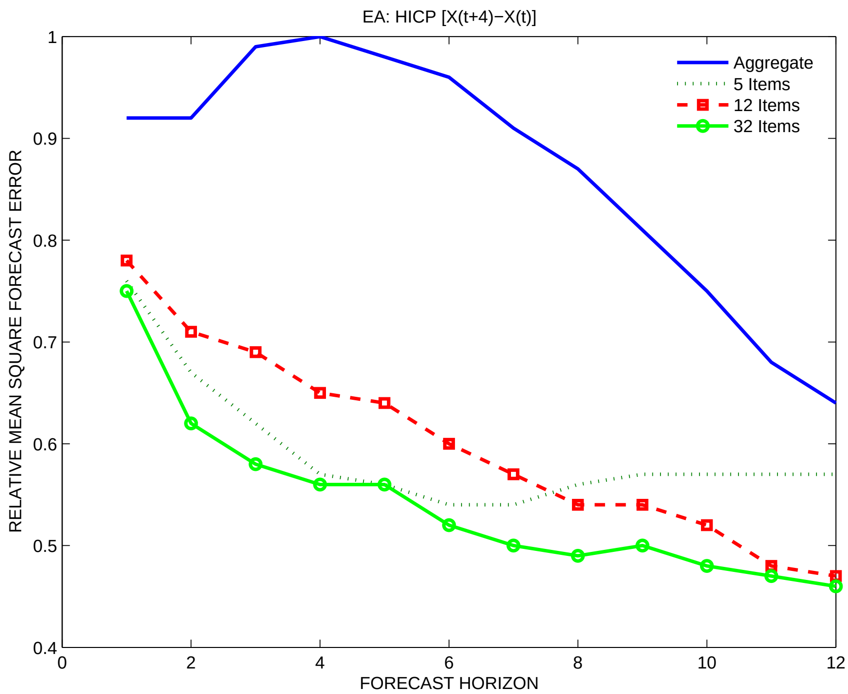


Figure 5: RMSFE of FAAR Model versus AO (US-YoY)

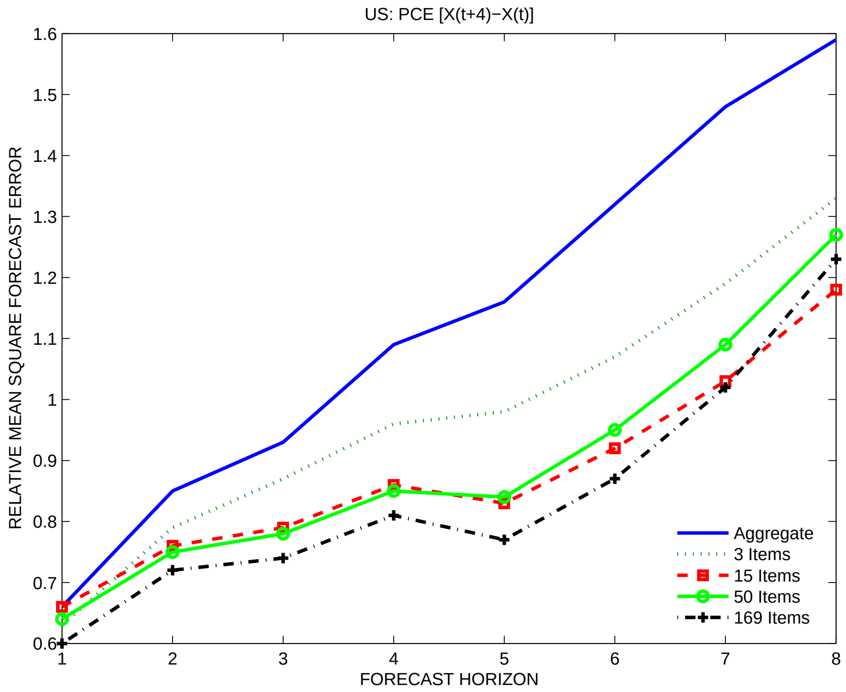


Figure 6: RMSFE of FAAR Model versus AO (EA-YoY)

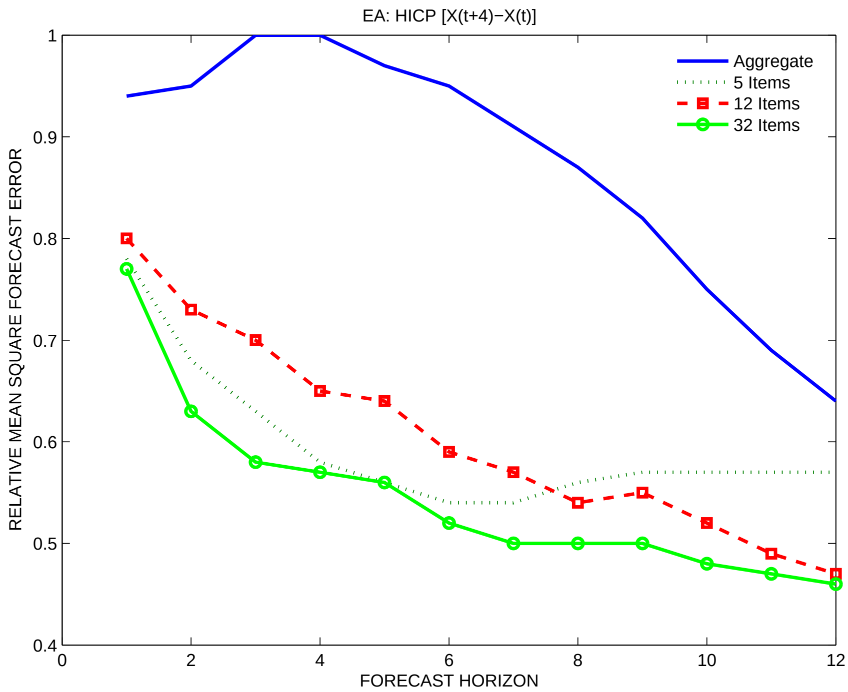


Figure 7: RMSFE of BVAR Model versus AO (US-YoY)

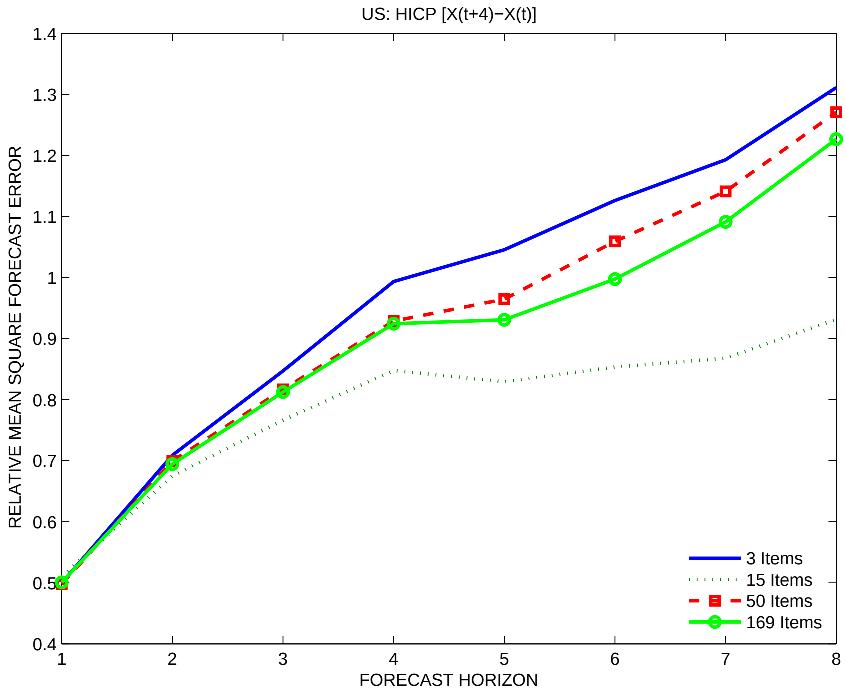


Figure 8: RMSFE of BVAR Model versus AO (EA-YoY)

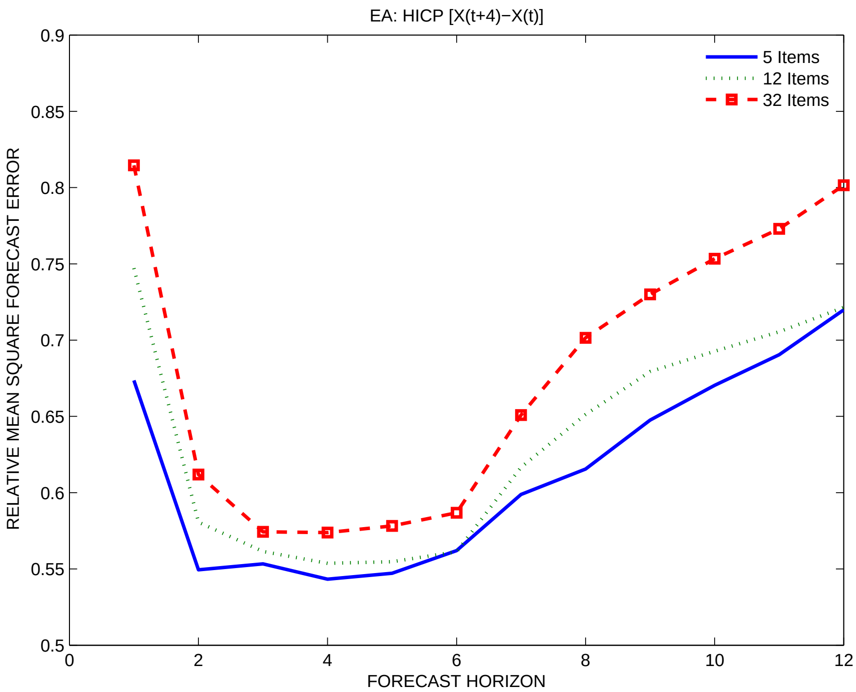


Figure 9: RMSFE of AR Model versus AO (US - h period inflation)

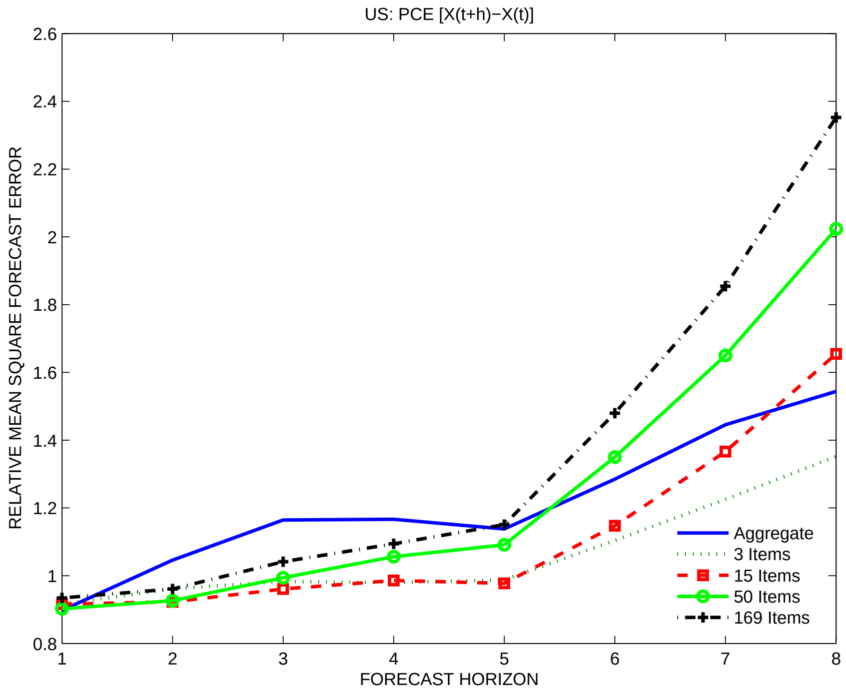


Figure 10: RMSFE of FAAR Model versus AO (US - h period inflation)

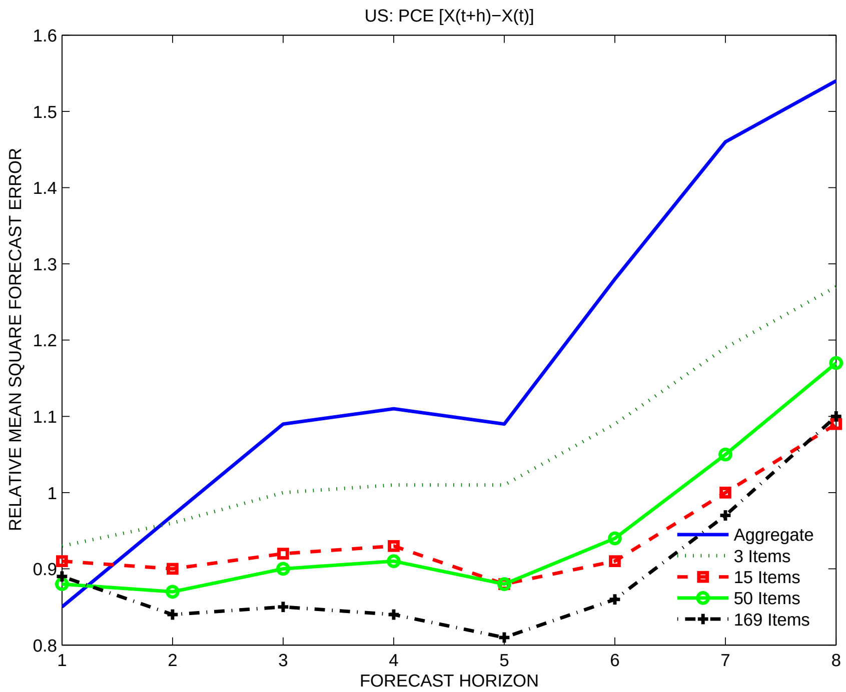


Figure 11: RMSFE of BVAR Model versus AO (US - h period inflation)

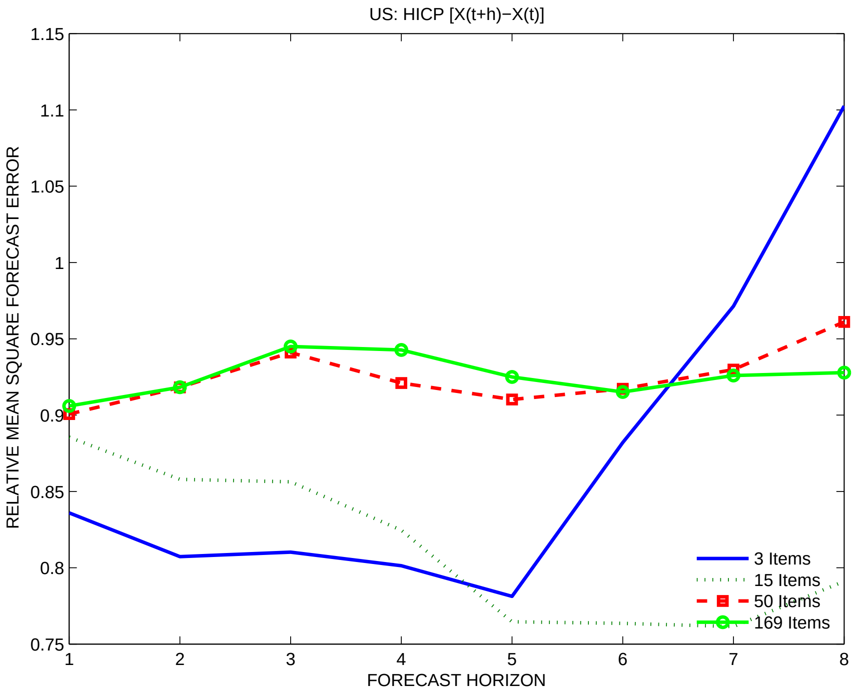
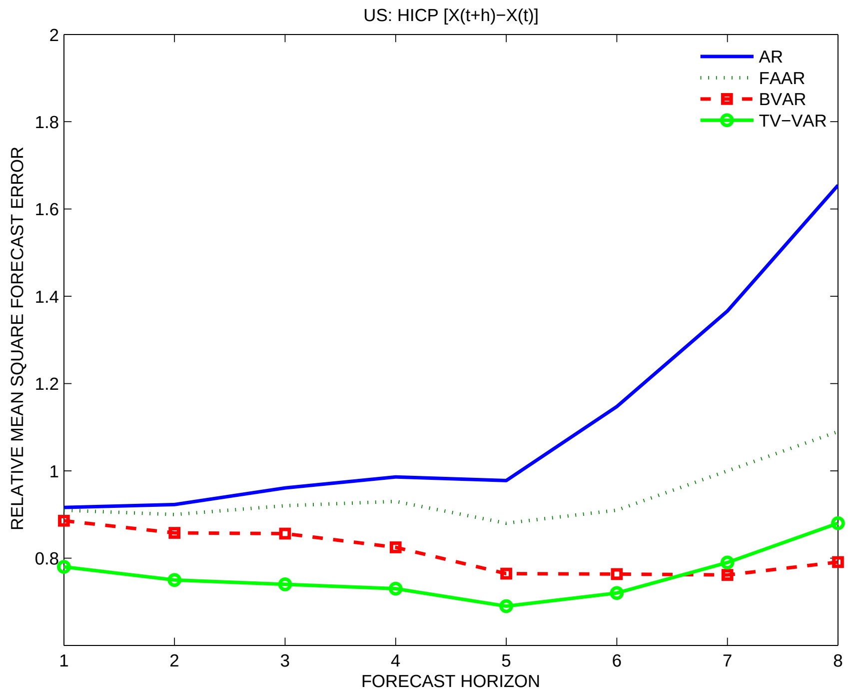


Figure 12: RMSFE of the AR, FAAR, BVAR and TV-AR Models versus Atkeson-Ohanian (US - h period inflation, 15 items)



Bibliography

[1] Al-Gudhea, S., T. Kenc and S. Dibooglu (2007), "Do retail gasoline prices rise more readily then they fall? A threshold cointegration approach", *Journal of Economics and Business*, Vol. 59, pp. 560-574.

[2] Andrews, D. (2001), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, Vol. 59, No.3, pp. 817-858.

[3] Asplund, M., R. Eriksson and R. Friberg (2000), "Price adjustments by a gasoline retail chain", *Scandinavian Journal of Economics*, Vol. 102, pp. 101-121.

[4] Atkeson, A. and L. Ohanian, (2001), "Are Phillips Curves Useful for Forecasting Inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 25, No. 1, Winter 2001, pp. 211.

[5] Bacon, R. W., (1991), "Rcokets and feathers: the asymmetric speed of adjustment of UK retail gasoline prices to cost changes", *Energy Economics*, Vol 13, pp. 211 - 218.

[6] Banbura M., D. Giannone and L. Reichlin (2010): "Large Bayesian VARs", Journal of Applied Econometrics, Vol. 25(1), pages 71-92,.

[7] Baxter, M. and R. King, (1999), "Measuring Business Cycle: Approximate Band-pass Filter for Economic Time Series", *The Review of Economics and Statistics*, Vol. 81, pp. 575 - 593.

[8] Benalal, N., Diaz del Hoyo, J. L., Landau, B., Roma, M. and Skudelny, F. (2004), "To aggregate or not to aggregate? euro area inflation forecasting", *ECB Working Paper 374*.

[9] Bermingham, C. (2007), "How Useful is Core Infation for Forecasting Headline Inflation?", *Economic and Social Review*, Vol. 38, No.3, pp. 355-377.

[10] Bilke, L. and L. Stracca (2007), "A persistence-weighted measure of core inflation in the Euro Area", *Economic Modelling*, Vol. 24, pp. 1032 - 1047.

[11] Bjornland, H. C. (2001), "Identifying Domestic and Imported Core Inflation", *Applied Economics*, Vol. 33, pp. 1819 - 1831.

[12] Blanchard, O. J. and D. Quah (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 79, pp. 655-673.

[13] Blinder, A. (1997), Commentary, *Federal Reserve Bank of St. Louis Economic Review*, May/June, pp 157-160.

[14] Borenstein, S., C.A. Cameron and R. Gilbert (1997), "Do Gasoline Prices Respond Asymetrically to Crude Oil Price Changes?", *Quarterly Journal of Economics*, Vol. 112, February, pp. 305-339.

[15] Breusch, T. (1978), "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economics Papers*, Vol. 17, pp. 334-355.

[16] Bruneau, C., De Bandt, O., Flageollet, A. Michaux, E. (2007), "Forecasting inflation using economic indicators: The case of France", *Journal of Forecasting*, Vol. 26, pp. 1 - 22.

[17] Bryan, M., S. Cecchetti, and R. Wiggins II (1997), "Efficient Inflation Estimation", *NBER Working Paper*, No. 6183.

[18] Carruth, A., M. Hooker and A. Oswald (1998), ".Unemployment Equilibria and Input Prices: Theory and Evidence from the United States", *Review of Economics and Statistics*, Vol. 80, pp. 621-628.

[19] Cecchetti, S. (1997), "Measuring Short-Run Inflation for Central Bankers", *Federal Reserve Bank of St. Louis Economic Review*, May/June, pp. 143-155.

[20] Chacra, M. (2002), "Oil Price Shocks and Retail Energy Prices in Canada", *Bank of Canada Working Paper*, 8.

[21] Chen, S. (2008), "Oil Price Pass-Through into Inflation", Available at SSRN: http://ssrn.com/abstract=1085941.

[22] Claus, I. (1997), "A Measure of Underlying Inflation in the United States", *Bank of Canada Working Paper 97-20*

[23] Clark, Todd. (2001), "Comparing Measures of Core Inflation", *Federal Reserve Bank of Kansas Economic Review*, 2nd Quarter.

[24] Clark T. and K. West, (2006), "Using out-of-sample mean squared prediciton errors to test the martingale difference hypothesis", *Journal of Econometrics*, Vol. 135, November, pp. 85-110.

[25] Clark T. and K. West, (2007), "Approximately normal tests for equal predictive accuracy in nested models", *Journal of Econometrics*, Vol. 138, November, pp. 155-186.

[26] Cogley, T. (2002), "A Simple Adaptive Measure of Core Inflation", *Journal of Money, Credit and Banking*, Vol. 34(1), pp. 94-113.

[27] Cogley, T., and A. M. Sbordone, (2008),"Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve", *American Economic Review*, Vol. 98(5), pp. 2101-26.

[28] Cutler, J. (2001), 'A New Measure of Core Inflation in the UK", *MPC Unit Discussion Paper*, Vol. 3.

[29] D'Agostino, A., Gambetti, L. and D. Giannone, (2009), "Macroeconomic Forecasting and Structural Change", *European Central Bank Working Paper No. 1167*.

[30] De Gregorio, J., O. Landerretche and C. Neilson, (2007), "Another Pass-Through Bites the Dust? Oil Prices and Inflation", *Central Bank of Chile Working Papers*, No. 417.

[31] Diebold F. and R. Mariano, (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-263.

[32] Dolmas, J. (2006), Commentary, *Federal Reserve Bank of Dallas Working Paper 0506*.

[33] Dow, J. (1994), "Measuring Core Inflation using Multiple Price Indexes", *Unpublished Manuscript*, Department of Economics, University of California-Riverside, June.

[34] Duarte, C. and Rua, A. (2007), "Forecasting inflation through a bottom-up approach: How bottom is bottom?", *Economic Modelling*, Vol. 24, No. 6, November 2007, pp. 941-953.

[35] Elliott, G., T. Rothenbery and J. Stock (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, Vol.64, No.4, July, pp.813-836.

[36] Geweke, J. (2004), "Issues in "rockets and feathers" gasoline price literature", Report to Federal Trade Commission.

[37] Godby, R., A.M. Lintner, T. Stengos and B. Wandschneider (2000), "Testing for asymmetric pricing in the Canadian retail gasoline market", *Energy Economics*, Vol. 22, pp. 349-368.

[38] Godfrey, L. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors include Lagged Dependent Variables", *Econometrica*, Vol. 46, pp. 1293-1302.

[39] Hamilton, J. (1983), "Oil and the Macroeconomy Since World War II", *Journal of Political Economy*, Vol. 91, pp. 228-248.

[40] Hamilton, J. (2008), "Oil and the Macroeconomy", *New Palgrave Dictionary of Economics*, Second Edition.

[41] Hendry, D. F. and Hubrich, K. (2006), "Forecasting aggregates by disaggregates", *European Central Bank Working Paper 589.*

[42] Hendry, D. F. and Hubrich, K. (2010), "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate", *Journal of Business and Economic Statistics.*

[43] Hogan, S., M. Johnson and T. Lafleche (2001), "Core Inflation", *Bank of Canada Technical Report No. 89.*

[44] Hooker, M. (2002), "Are Oil Shocks Inflationary? Asymmetric and Non-Linear Specifications versus Change in Regimes", *Journal of Money, Credit and Banking*, Vol. 34, No.2, pp. 540-561.

[45] Hubrich, K. (2005), "Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?", *International Journal of Forecasting*, Vol. 21, pp. 119-136.

[46] Kearns, J. (1998), "The Distribution and Measurement of Inflation", *Reserve Bank of Australia Discussion Paper 9810*

[47] Kenny, D. and D. McGettigan, (1996), "Non-Traded, Traded and Aggregate Inflation in Ireland: Further Evidence", Central Bank of Ireland, Technical Paper 5/RT/96

[48] Kenny, D. and D. McGettigan, (1996), "Exchange Rate Pass-Through and Irish Import Prices", Central Bank of Ireland, Technical Paper 6/RT/96

[49] Kohn, R. (1982), "When is an aggregate of a time series efficiently forecast by its past?", *Journal of Econometrics*, Vol. 18, pp. 337-349.

[50] Landau, B. (2000), "Core Inflation Rates: A Comparison of Methods based on West German Data",*Discussion Paper Series 1: Economic Studies*, 2000/04, Deutsche Bundesbank.

[51] Litterman, R.B., (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", *Journal of Business and Economic Statistics*, Vol.4, No.1, pp. 25-38.

[52] Lutkepohl, H. (1984a), "Forecasting contemporaneously aggregated vector ARMA processes", *Journal of Business and Economic Statistics*, Vol.2, No.3, pp. 201-214.

[53] Lutkepohl, H. (1984b), "Linear transformations of vector ARMA processes", *Journal of Econometrics*, Vol.26, pp. 283-293.

[54] Machlem, T (2001), “A new Measure of Core Inflation”, *Bank of Canada Review*, Autumn.

[55] Manera, G. and M. Manera (2007), “Econometric models of asymmetric price transmission”, *Journal of Economic Surveys*, Vol. 21, No. 2.

[56] Manning, D. N., (1991), “Petrol prices, oil price rises and oil price falls, some evidence for the UK since 1972”, *Applied Economics*, Vol. 23, pp. 1535-1541.

[57] Marcellino, M., Stock, J. H. and Watson, M. W. (2003), “Why Has US Inflation Become Harder to Forecast?”, *Journal of Money, Credit and Banking*, Vol. 39, pp. 3-33.

[58] McDermott, J. and A. Scott (1999), “Concordance in Business Cycles”, *Reserve Bank Discussion Paper*, G99/7, Reserve Bank of New Zealand.

[59] Meyler, A. (1999), “A Statistical Measure of Core Inflation”, *Central Bank of Ireland Research Technical Paper 2/RT/99*

[60] Meyler, A. (2010), “The pass-through of oil prices into consumer liquid fuel prices in an environment of high and volatile oil prices”, *Energy Economics*, Vol. 31, No. 6, November 2009, pp. 867-881.

[61] Moser, G., Rumler, F. Scharler, J. (2007), “Forecasting Austrian inflation”, *Economic Modelling*, Vol. 24, Issue 3, May 2007, pp. 470-480.

[62] National Consumer Agency (2008), “Investigation into Petrol and Diesel Price Movements”, Report.

[63] Ng, S. and P. Perron (2001), “Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power”, *Econometrica*, Vol.69, No.6, November, pp.1519-1554.

[64] P. Perron (1989), “The Great Crash, the Oil Price Shock and the Unit Root Hypothesis”, *Econometrica*, Vol.57, No.6, November, pp. 1361-1401.

[65] Peseran, M. H., Y. Shin, and R.J. Smith (2001), “Bounds Testing Approaches to the Analysis of Level Relationships”, *Journal of Applied Econometrics*, Vol. 18, pp. 281-295.

[66] Quah, D. and S. Vahey (1995), “Measuring Core Inflation”, *The Economic Journal*, Vol. 105, No. 432, pp. 1130 - 1144.

[67] Rich, R. and C. Steindel (2007), “A Comparison of Measures of Core Inflation”, *Economic Policy Review*, Federal Reserve Bank of New York, December, pp. 19-38.

[68] Roger, S. (1998), “Core Inflation: Concepts, Uses and Measurement”, *Reserve Bank of New Zealand Discussion Paper Series G98/9*

[69] Rose, D.E. (1977), "Forecasting aggregates of independent ARIMA processes", *Journal of Econometrics* Vol. 5, pp.323-345.

[70] Rotemberg, J. and M. Woodford, (1996), "Imperfect Competition and the Effects of Energy Price Increases", *Journal of Money, Credit, and Banking*, Vol. 28, pp. 549-577.

[71] Smith J. (2004), "Weighted Median Inflation: Is This Core Inflation", *Journal of Money, Credit and Banking*, Vol. 36(2), pp. 253-263.

[72] Stock, J. M. Watson (2002), "Macroeconomic Forecasting using Diffusion Indexes", *Journal of Business and Economic Statistics*, Vol. 20(2), pp. 147-162.

[73] Stock, J. H. and Watson, M. W. (2003), "Macroeconomic forecasting in the euro area: Country specific versus area-wide information", *European Economic Review*, Vol. 47, pp. 1-18.

[74] Stock, J. and M. Watson (2007), "Why Has US Inflation Become Harder to Forecast?", *Journal of Money, Credit and Banking*, Vol. 39, No. 1, February 2007.

[75] Slevin, G., (2003), "Structural Model of Irish Inflation", Central Bank of Ireland, Technical Paper 1/RT/03

[76] Slevin, G., (2003), "Potential Output and the Output Gap in Ireland", Central Bank of Ireland, Technical Paper 5/RT/01

[77] Tiao, G. C. and Guttman, I. (1980), "Forecasting contemporal aggregates of multiple time series", *Journal of Econometrics*, Vol. 12, pp. 219-230.

[78] Wynne, M. (1999), "Core Inflation: A Review of Some Conceptual Issues", *European Central Bank Working Paper*, No. 5.

[79] Zellner, A. and Tobias, J. (2000), "A note on aggregation, disaggregation and forecasting performance", *Journal of Forecasting* Vol. 19, pp. 457-469.